



UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS

FACULTAD DE NEGOCIOS

PROGRAMA ACADÉMICO DE ADMINISTRACIÓN DE EMPRESAS

Predecir la demanda de alumnos matriculados en el CCDG para el año 2020

TRABAJO DE INVESTIGACIÓN

Para optar al grado de Bachiller en Administración de Empresas

AUTOR(ES)

Bravo España, Fiorella Aracely (<https://orcid.org/0000-0002-0674-9237>)

Chávarri Aguilar, Maite Yanín (<https://orcid.org/0000-0003-0204-0578>)

Huayta León, María de los Ángeles (<https://orcid.org/0000-0001-6749-0188>)

ASESOR(ES)

Palacios Ruiz, Julio César (<https://orcid.org/0000-0003-1610-5205>)

Lima, 9 de diciembre de 2020

DEDICATORIA

El presente trabajo lo dedicamos a nuestros padres, pues gracias a su apoyo moral estamos a un paso de concluir con una etapa de nuestra vida, esta es, terminar nuestra carrera profesional. Ellos fueron nuestros principales soportes durante este largo proceso dentro de la universidad, y gracias a los valores y principios que nos inculcaron desde pequeñas, pudimos ser constantes y perseverantes para poder cumplir con este objetivo.

AGRADECIMIENTOS

El agradecimiento del presente proyecto va dirigido, en primer lugar, a Dios, por permitirnos culminarlo con éxito. También al docente del curso Julio Palacios, pues gracias a sus conocimientos, cubrimos las expectativas del trabajo final. Finalmente, a nuestro contacto dentro de la empresa Centro de Capacitación y Desarrollo Global, Edith Rivera, quien nos facilitó la data estudiada y analizada en el trabajo.

RESUMEN

El presente trabajo de investigación tuvo como objetivo determinar la cantidad de alumnos matriculados en los diversos programas para el año 2020 dentro del Centro de Capacitación y Desarrollo Global (CCDG), ubicado en la ciudad de Lima; ello, a través del análisis del historial de la data de los años 2018 y 2019. Se ejecutó una investigación predictiva, dado que dicho tipo de investigación tiene como propósito prever o anticipar situaciones futuras, con el fin de tomar buenas decisiones en cuanto a la apertura de vacantes de los diversos programas y diplomados para cubrir la posible demanda de alumnos matriculados y, por otro lado, determinar las variables que permiten captar un mayor número de clientes.

La metodología de ciencia de datos aplicada fue la regresión lineal simple; esta es una técnica que ayuda a proyectar el número de alumnos que podrían matricularse en el periodo 2020. Las variables utilizadas para esta técnica fueron la cantidad de alumnos matriculados y el número de meses correspondientes a los dos últimos años analizados.

Como resultado, se obtuvo que, posiblemente, se tendrían 3280 alumnos matriculados para el año 2020 en los programas del CCDG, y que la variable más influyente para la atracción son las clases de modalidad virtual; asimismo, los cursos más demandados son la certificación del Organismo Supervisor de las Contrataciones del Estado (OSCE) y los diplomas del Sistema Integrado de Administración Financiera (SIAF). Los medios de contacto más efectivos, y con los que es posible atraer un mayor número de clientes, son la aplicación de WhatsApp y las llamadas telefónicas.

Palabras clave: regresión lineal, medio de contacto, metodología, captación de clientes, diplomados, cursos, programas, demanda, análisis.

ABSTRACT

The present research work aims to determine the number of students enrolled in the various programs for the year 2020 within the Global Training and Development Center, located in the city of Lima. This, through the analysis of 2018 and 2019 data history. A predictive research was carried out, since this type of research is intended to foresee or anticipate future situations, in order to make good decisions regarding the opening of vacancies in the various programs and diplomas to cover the possible demand of enrolled students, and, also, to determine the variables that allow a greater customer acquisition.

The applied Data Science methodology is simple linear regression, which is a technique that projects the number of students who would enroll in the 2020 period. The variables used for this technique were the number of enrolled students and the number of months corresponding to the last two years analyzed.

As a result, there would possibly be 3280 students enrolled by 2020 in the programs offered at the CCDG, and the most influential variable for customer recruitment are virtual classes; also, the most demanded course is OSCE certification and SIAF diplomas. The most effective means of contact, which can attract more clients, are WhatsApp and phone calls.

Keywords: linear regression, means of contact, methodology, customer acquisition, diplomas, courses, programs, demand, analysis.

ÍNDICE DE CONTENIDO

1	INTRODUCCIÓN.....	11
2	PRESENTACIÓN DE LA EMPRESA PARA EL PROYECTO	14
2.1	Presentación y descripción de la empresa.....	14
2.1.1	Misión.....	14
2.1.2	Visión	15
2.1.3	Objetivos de la empresa.....	15
2.2	Modelo de negocio (Canvas)	15
2.2.1	Ventaja competitiva.....	16
2.3	Análisis PESTEL	16
3	COMENZANDO EL PROYECTO.....	20
3.1	Identificación del problema	20
3.2	Formulación de preguntas para Data Science	21
3.3	Justificación	21
3.4	Objetivos	22
3.4.1	Objetivo general	22
3.4.2	Objetivos específicos.....	22
3.5	Hipótesis	22
4	MARCO TEÓRICO	23
4.1	Metodología de ciencia de datos.....	23
4.2	Gobierno de datos	27
4.3	Diccionario de datos	29
5	ARQUITECTURA DE DATOS	29
5.1	Arquitectura de datos funcional.....	29
	Arquitectura de datos tecnológica	30
5.2	30

5.3	Comprensión y preparación de los datos	33
5.3.1	Revisión y depuración de datos que se usaron en el estudio	33
5.4	Requisitos y recopilación de datos.....	34
5.4.1	Determinación de datos necesarios para responder a las preguntas	34
5.5	Ubicación de fuentes de datos	35
5.6	Presentación de datos recopilados	36
5.7	Uso de herramientas para determinar posibles relaciones entre los datos	38
5.7.1	Determinando el punto de partida	38
5.7.2	Determinación de las herramientas de visualización a utilizar para el proyecto	39
6	PRODUCCIÓN Y ANÁLISIS EXPLORATORIO DE DATOS.....	39
6.1	Plan de muestreo y tamaño de la muestra.....	39
6.1.1	Definir la población objetivo.....	39
6.1.2	Identificar las unidades de prueba	40
6.1.3	Seleccionar la técnica de muestreo.....	40
6.2	Tamaño de la muestra	40
6.3	Diseño del estudio de muestreo	41
6.4	Ejecución del muestreo	41
6.5	Resumen y examen de la distribución de variables	42
6.6	Distribución de la muestra: interpretación del comportamiento.....	43
6.7	Presentación de datos a utilizar para el estudio	43
6.8	Resultado del análisis de relaciones.....	49
6.9	Estudio probabilístico de los datos obtenidos.....	51
6.9.1	Técnica de muestreo probabilístico	51
7	TÉCNICAS DE CIENCIA DE DATOS	54
7.1	El modelo y la historia	57

7.1.1	Determinación del modelo más apropiado para la descripción o predicción.	57
8	DESCRIPCIÓN DE LOS HALLAZGOS	59
8.1	Visualización, implementación y ajustes.....	59
8.1.1	Definición y justificación de los criterios a tomar para la visualización de los resultados	59
8.2	Visualización de los hallazgos	60
8.3	Estrategia de implementación	63
8.4	Calibrado e implementación	64
8.4.1	Métodos de ajuste	64
8.5	Aplicando la calibración del modelo	67
9	Retroalimentación.....	68
	Calibrando la regresión calculada.....	70
	9.1.....	70
10	CONCLUSIONES.....	72
11	RECOMENDACIONES	73
12	FACTORES CRÍTICOS DE ÉXITO PARA EL PROYECTO	74
13	NUEVAS PREGUNTAS	75
14	DECISIONES A TOMAR A PARTIR DE LOS DATOS ESTUDIADOS	76
15	BIBLIOGRAFÍA	77
16	ANEXOS	79

ÍNDICE DE TABLAS

Tabla 1. Factores externos	17
Tabla 2. FODA	18
Tabla 3. Diccionario de datos	29
Tabla 4. Fuentes de datos	35
Tabla 5. Datos con estadística	52
Tabla 6. Proyección 2020	55

ÍNDICE DE FIGURAS

Figura 1. Valores	11
Figura 2. Regresión lineal de ventas por mes	12

Figura 3. Modelo de negocios - Canvas	16
Figura 4. Ingresos por años	21
Figura 5. Metodología de datos	24
Figura 6. Arquitectura de datos funcional	30
Figura 7. Arquitectura de datos tecnológica.....	31
Figura 8. Power BI	33
Figura 9. Seace: nuevos procesos publicados por día	37
Figura 10. Contrataciones 2017-2018	38
Figura 11. Estado de procesos de contratación 2018.....	38
Figura 12. Power BI	39
Figura 13. Cálculo de la muestra.....	41
Figura 14. <i>Dashboard</i> de desarrollo global	44
Figura 15. Ingreso por clase	45
Figura 16. Edad promedio de estudiantes.....	45
Figura 17. Porcentaje por modalidad de estudio	46
Figura 18. Monto por programa	47
Figura 19. Monto por medio de contacto	48
Figura 20. Monto por vendedor.....	49
Figura 21. Relación de variables de modalidad virtual y programa.....	50
Figura 22. Relación de variables modalidad presencial con programa	50
Figura 23. Relación entre la variable clase y programa	51
Figura 24. Datos para análisis	53
Figura 25. Visualización de hallazgos.....	61
Figura 26. Monto por vendedor y medio de contacto	62

Figura 27. Monto por nombre del programa	62
Figura 28. Monto por mes y programa	63
Figura 29. Demanda proyectada	65
Figura 30. Demanda real al 40 %	66
Figura 31. Margen de error	66
Figura 32. Calibración	67
Figura 33. Demanda proyectada al 70 %	68
Figura 34. Demanda real al 30 %	69
Figura 35. Margen de error	70
Figura 36. Calibración ajustada	70

1 INTRODUCCIÓN

La ciencia avanza a pasos agigantados cada día; esto puede llevar a resultados y cambios inimaginables. Sin embargo, el cambio y la revolución tecnológica de los datos vienen de la mano, y es gracias al Data Science que en la actualidad se puede medir su influencia en el mundo. Existe una frase muy conocida por todos que señala que “el futuro es de los datos”; de hecho, se sabe que actualmente se producen dos millones y medio de terabytes de datos por día, dentro de los cuales se incluyen los de las redes sociales, como WhatsApp, las famosas búsquedas en Google, los tuits y los correos. A simple vista, parecen no tener ningún valor; no obstante, son invaluable en conjunto, debido al volumen que se genera por minuto.

FECHA	AÑO	MEDIO DE PAGO	MONTO	PROGRAMA	MODALIDAD	VENDEDOR	MEDIO DE CONTACTO
2/01/2018	2018	PAYU	257.00	CURSO PREPARATORIO OSCE II	VIRTUAL	DANIELA	ZOPIM
2/01/2018	2018	PAYU	457.00	CURSO PREPARATORIO OSCE	PRESENCIAL	JELISSA	WHATSAPP
2/01/2018	2018	BGP	457.00	CURSO SIAF BASICO	PRESENCIAL	JELISSA	WHATSAPP
2/01/2018	2018	BN	257.00	CURSO PREPARATORIO OSCE II	VIRTUAL	MARISOL	FACEBOOK
2/01/2018	2018	BGP	257.00	CURSO OSCE II	VIRTUAL	MARISOL	WEB
2/01/2018	2018	INTERBANK	257.00	CURSO OSCE II	VIRTUAL	MARISOL	WEB
3/01/2018	2018	BN	257.00	CURSO PREPARATORIO OSCE II	VIRTUAL	DANIELA	ZOPIM
3/01/2018	2018	BN	248.50	CURSO OSCE II	VIRTUAL	DANIELA	ZOPIM
3/01/2018	2018	BGP	397.00	CERTIFICADO INVIERTE PERÚ	VIRTUAL	DANIELA	FACEBOOK
3/01/2018	2018	BN	141.50	CURSO PREPARATORIO OSCE II	VIRTUAL	MARISOL	WEB
3/01/2018	2018	PAYU	257.00	CURSO PREPARATORIO OSCE II	VIRTUAL	MARISOL	WEB
3/01/2018	2018	VISA	457.00	CURSO PREPARATORIO OSCE II	PRESENCIAL	JELISSA	LLAMADA
3/01/2018	2018	BGP	257.00	CURSO PREPARATORIO OSCE II	VIRTUAL	MARISOL	LLAMADA
4/01/2018	2018	BGP	397.00	CERTIFICADO INVIERTE PERÚ	VIRTUAL	DANIELA	WEB
4/01/2018	2018	BN	1297.00	DIPLOMADO GESTIO PUBLICA	VIRTUAL	MARISOL	WEB
4/01/2018	2018	BN	1067.00	DIPLOMA SIAF SIGA SEACE	VIRTUAL	JELISSA	LLAMADA
4/01/2018	2018	BGP	257.00	CURSO PREPARATORIO OSCE II	VIRTUAL	DANIELA	ZOPIM
4/01/2018	2018	BGP	1257.00	DIPLOMA SIAF SIGA SEACE	PRESENCIAL	JELISSA	OFICINA
4/01/2018	2018	BGP	480.00	CURSO PREPARATORIO OSCE II	VIRTUAL	MARISOL	WEB
4/01/2018	2018	INTERBANK	257.00	CURSO PREPARATORIO OSCE II	VIRTUAL	DANIELA	LLAMADA
4/01/2018	2018	BGP	370.00	DIPLOMA SIAF II (BÁSICO E INTERMEDI)	VIRTUAL	DANIELA	LLAMADA
4/01/2018	2018	BGP	417.00	DIPLOMA SIAF SIGA Y SEACE II	VIRTUAL	JELISSA	ZOPIM
4/01/2018	2018	BGP	157.00	CURSO PREPARATORIO OSCE II	VIRTUAL	MARISOL	LLAMADA
5/01/2018	2018	VISA	457.00	SIAF BASICO	PRESENCIAL	JELISSA	OFICINA

Figura 1. Valores

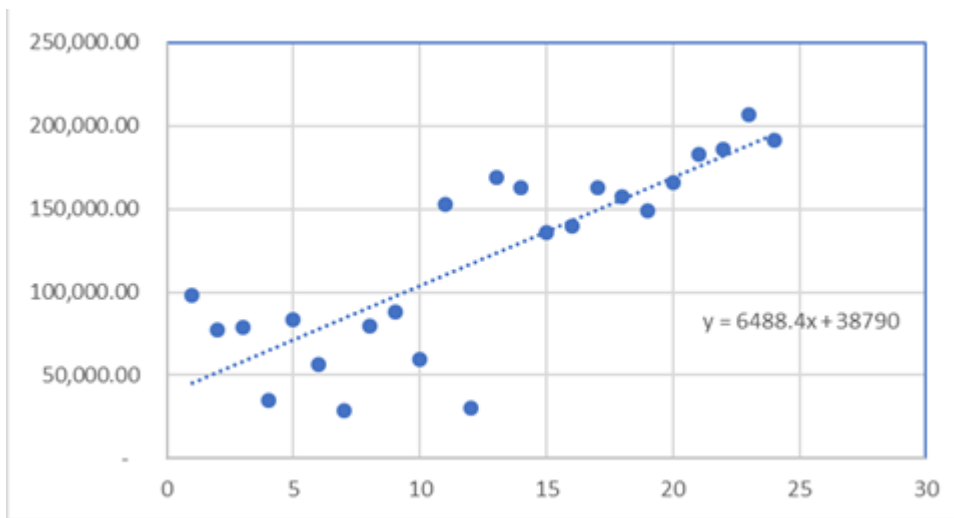


Figura 2. Regresión lineal de ventas por mes

Fuente: elaboración propia

Para abordar el análisis de este activo, fue que surgió hace unos años la ciencia de datos, la cual se encarga precisamente de tratar toda esa información que se genera, almacenarla y darle un sentido. Información de quiénes somos, qué hacemos, qué nos gusta e incluso, qué nos gustaría hacer en el futuro. Cada vez hay más información nuestra en internet (lo que se denomina huella digital) y por esta razón el volumen de datos ha ido creciendo exponencialmente con el paso de los años. (Universidad de Alcalá, s.f., párr. 3)

Para el presente proyecto, se aplicaron diversas herramientas de Data Science, como Power BI y Excel; estas contribuyeron a la construcción de gráficos a modo de *dashboard*, a fin de profundizar el análisis de las variables dentro de la data extraída de la misma fuente de la empresa. Ello, en función de un objetivo general y unos objetivos específicos, los cuales dan paso a la formulación de preguntas que fueron resueltas al finalizar el presente trabajo; todo esto se ejecutó con la data del CCDG.

Esta es una empresa educativa privada, dedicada a proveer desarrollo gerencial de amplio rango a los servidores privados y públicos, así como a técnicos y profesionales del país, a

través de los servicios educativos en temas de gestión. La empresa fue creada en el año 2011 y desde entonces cuenta con programas en línea y de manera presencial, como diplomas, cursos, seminarios, conferencias y talleres, con el objetivo de aportar a sus alumnos una educación de calidad y con la mejor plana docente.

2 PRESENTACIÓN DE LA EMPRESA PARA EL PROYECTO

2.1 Presentación y descripción de la empresa

El Centro de Capacitación y Desarrollo Global EIRL (Empresa Individual de Responsabilidad Limitada), con RUC (Registro Único de Contribuyentes) 20544396791, es una empresa privada de origen individual que cuenta con nueve años en el mercado peruano, y sus ventas ascienden a dos millones de soles anuales. La compañía inició sus operaciones en el año 2011, con la idea de cubrir un nicho de mercado identificado; esta decidió proveer servicios de capacitación para los trabajadores administrativos del Estado en la ciudad de Lima o en provincias del país.

La empresa cuenta con la certificación ISO 9001 2015, el cual es un estándar internacional de la ISO (Organización Internacional de Normalización) para las entidades que mejoran el rendimiento general e implementan un sistema de gestión de calidad para demostrar los requisitos de calidad para sus actividades. La mejora continua de dichos procesos garantiza el crecimiento sostenido del negocio, lo que lo lleva a mantener una relación mutuamente beneficiosa con los elementos en su entorno. En tal sentido, esta proporciona alternativas de solución con los recursos apropiados para mejorar los servicios con más eficiencia y veracidad. La compañía, además, ha formalizado convenios con las siguientes universidades:

2.1.1 Misión

Desarrollar y fortalecer las competencias profesionales de los funcionarios públicos y privados para impulsar el talento y la productividad.

2.1.2 Visión

Ser una empresa educativa estratégica de las organizaciones públicas y privadas para el desarrollo de su talento humano.

2.1.3 Objetivos de la empresa

- Brindar un servicio de calidad, ayudando a la formación profesional de nuestros alumnos con excelencia.
- Tener una plana docente capacitada, con un buen nivel de enseñanza y experiencia pedagógica.
- Estar dentro de los centros de capacitación con mayor prestigio y mayor inserción laboral de nuestros egresados.

2.2 Modelo de negocio (Canvas)

La empresa está orientada a brindar servicios de capacitación, como cursos, diplomas y diplomados a nivel nacional, y se especializa en la gestión pública. Asimismo, ha desarrollado una plataforma única para la realización de las clases en línea, lo cual mejora la calidad del trabajo. Por otro lado, cuenta con seis salas de capacitación y personal capacitado.

Figura 3. Modelo de negocios - Canvas



Fuente: elaboración propia

2.2.1 Ventaja competitiva

La ventaja competitiva que tiene la empresa es que cuenta con certificaciones vigentes de calidad de enseñanza y servicio a disposición del estudiante. El sistema de gestión de calidad se basa en la norma ISO 9001, pues dicha certificación se obtuvo para garantizar a los clientes la eficacia de los productos o servicios. Por esto, el cliente potencial que decide crecer profesionalmente tiene a su alcance una enseñanza de calidad, con una inversión que justifica una mejor clasificación en el mercado laboral. Esto brinda al egresado una ventaja de formación sobre los sistemas actuales que utiliza el gobierno, a fin de competir por plazas públicas y privadas; además, cuenta con convenios con tres universidades de prestigio.

2.3 Análisis PESTEL

“En teoría, el análisis PESTEL es un instrumento de planificación estratégica para definir el contexto de una campaña. Analiza factores externos como son los políticos, económicos,

sociales, tecnológicos, ambientales y jurídicos los cuales pueden influir en la campaña” (ONU Mujeres, 2012, párr. 1). Al realizar dicho análisis, es necesario describir el contexto relacionado con la empresa y cómo se vería afectada en relación con las variables de la herramienta.

Tabla 1. Factores externos

FACTORES POLÍTICOS	FACTORES ECONÓMICOS
<p>P - La crisis gubernamental actual en el país afecta los proyectos relacionados con el emprendimiento.</p> <ul style="list-style-type: none"> - Reducción de la cobertura social pública. 	<p>E – Actualmente, debido al COVID-19, existe una crisis económica que afecta tanto a las empresas como a las personas naturales.</p> <ul style="list-style-type: none"> - Limitación en los presupuestos públicos. - Costos de equipamiento, dado que los cursos son virtuales. - Desempleo y cierre de empresas.
FACTORES SOCIALES	FACTORES TECNOLÓGICOS
<p>S - La escasez de movilidad de las personas debido a la pandemia del COVID-19.</p> <ul style="list-style-type: none"> - Aspectos éticos y los valores de la sociedad. - Alfabetización digital y accesibilidad a internet de todos los grupos etarios. 	<p>T - Adaptación y desarrollo de las nuevas tecnologías al alcance.</p> <ul style="list-style-type: none"> - Difusión del aprendizaje autónomo y de la modalidad a distancia. - Crecimiento del acceso de la población a internet y a las comunicaciones móviles. - Protección de datos y seguridad.
FACTORES ECOLÓGICOS	FACTORES LEGALES

E - Políticas de protección medioambiental y las leyes que influyen en ella.

- Sociedad que valora mucho la empresa con responsabilidad social corporativa (RSC); políticas para reducir el impacto ambiental que pueden mejorar la imagen de la empresa.

L - Cambio en la reglamentación de la Superintendencia Nacional de Educación Superior Universitaria (Sunedu).

- El sector de la actividad está adecuadamente regulado y protegido.
- Licencia actualizada y en funcionamiento.
- Cumplir con los requisitos para el otorgamiento de los certificados.

Fuente: elaboración propia

Otras herramientas de análisis

Tabla 2. FODA

Fortalezas	Oportunidades
<ul style="list-style-type: none"> • Convenios con la Universidad Nacional Mayor de San Marcos (UNMSM), la Universidad Nacional del Callao (UNAC) y la Universidad Nacional José Faustino Sánchez Carrión (UNJFSC). • Certificación ISO 9001 2015. • Enseñanza con <i>software</i> actualizado. • Plataforma interactiva única. 	<ul style="list-style-type: none"> • Servidores públicos que necesitan ser promovidos. • Demanda de nuevo personal en la gestión pública. • Acceso a internet en más provincias. • Cambios constantes en las leyes o reglamentos del Estado.

-
- Asistencia personalizada.
 - Docentes especializados.

Debilidades	Amenazas
<ul style="list-style-type: none"> • Falta de mecanismos de control de calidad de temas dictados. • Falta de publicidad para dar a conocer las fortalezas de la empresa. • Falta de promoción y venta de los cursos <i>in-house</i>. • Presencia reducida en el mercado. • Rotación de personal. 	<ul style="list-style-type: none"> • Gran oferta de servicios similares en el mercado. • Surgimiento de nuevas empresas. • Servicios similares con precios muy bajos. • Desastres naturales que afectan a los servicios de energía, internet y otros. • Deficiencias en el servicio de internet del proveedor.

Fuente: elaboración propia

Principales competidores:

- **Instituto Peruano de Asuntos Públicos, Política y Gobierno (IPAPPG):** es una organización privada con presencia nacional dedicada a brindar servicios de capacitación, asesoría y consultoría a profesionales y organizaciones, con el objetivo de hacerlos más eficientes y productivos en sus actividades laborales. Para brindar programas de excelente calidad cuentan con la certificación de la Universidad Nacional Mayor de San Marcos mediante distintos convenios y una plana docente compuesta por profesionales de alto nivel de ministerios (MEF, Minsa, PCM, etc.). (IPAPPG, s.f., párr. 1)
- **Centro Peruano de Estudios Gubernamentales (CEPEG):** es una asociación civil sin fines de lucro, dedicada a la enseñanza en temas relacionados con la administración pública, las ciencias sociales y jurídicas en los distintos ámbitos gubernamentales del país y sus instituciones. Las actividades académicas que realizan buscan contribuir al mejoramiento de la función pública, acrecentando así los conocimientos, habilidades y aptitudes de los funcionarios y servidores de las organizaciones públicas y privadas. En cumplimiento de su misión institucional y promoviendo la pluralidad de enfoques interdisciplinarios, realizan proyectos de investigación y consultorías en diversos campos cuyos resultados de dichas investigaciones

serán difundidas a través de su anhelante labor editorial. (CEPEG, s.f., párr. 1)

- **R&C Consulting:** contiene lo último en gestión pública y los sistemas administrativos del Estado, ya que son herramientas de uso obligatorio para trabajar en las instituciones del Estado, o requisitos mínimos para postular a los puestos administrativos como CAS. (R&C Consulting, s.f., párr. 1)

Todos los cursos incluyen una metodología basada en el desarrollo de competencias materiales de estudio. Debido al crecimiento de la compañía, esta actualmente cuenta con 30 colaboradores, incluyendo al gerente, el administrador, el contador, el asistente de gerencia y el jefe de marketing.

3 COMENZANDO EL PROYECTO

3.1 Identificación del problema

Se identifican los datos relevantes de los alumnos que se matriculan en los cursos, diplomas y diplomados, como los pagos realizados y el medio de contacto, pero la empresa muchas veces se queda sin poder cubrir la cantidad demandada de potenciales alumnos matriculados. Por ello, se buscó predecir la demanda para el presente año, con el fin de planificar con mayor precisión las vacantes para el año en curso 2020.

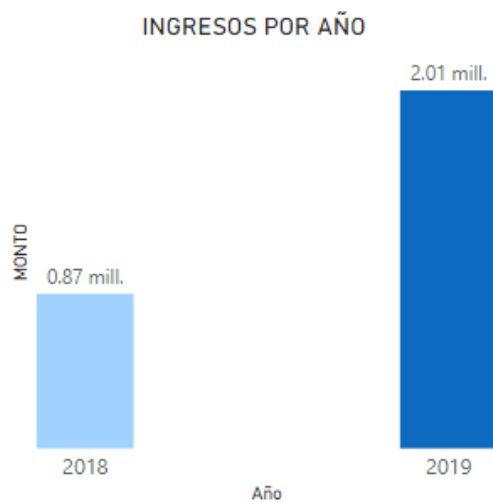


Figura 4. Ingresos por años

Fuente: elaboración propia

3.2 Formulación de preguntas para Data Science

- ¿Cuál es el medio de contacto más efectivo para cerrar la venta?
- ¿Cuál sería la demanda de alumnos proyectada para el 2020?
- ¿Cuál es la modalidad de enseñanza que prefieren los alumnos?

3.3 Justificación

El propósito de este trabajo de investigación es identificar las variables que influyen en los alumnos para su matrícula en el CCDG, debido a que este es el medio de contacto más eficiente para atraer nuevos alumnos. Además, es importante determinar cuál es la modalidad de enseñanza que prefieren los alumnos, teniendo en cuenta la coyuntura actual de la sociedad mundial. Asimismo, se busca tener una proyección de la demanda más exacta en cuanto a los posibles alumnos que requieran los servicios de capacitación. Por último, se propone brindar educación a las personas desempleadas, de manera que puedan insertarse en nuevos puestos

del Estado peruano o en empresas privadas, con una remuneración más alta que la promedio, gracias a los nuevos conocimientos obtenidos dentro de la institución.

3.4 **Objetivos**

3.4.1 **Objetivo general**

- Predecir la demanda de posibles alumnos a matricularse para el 2020 y las variables que influyen en la captación de potenciales ventas de los servicios de capacitación.

3.4.2 **Objetivos específicos**

- Conocer cuál es el medio de contacto más adecuado para contactar más clientes.
- Identificar el curso que tendría una mayor demanda entre los alumnos.
- Identificar las preferencias de los alumnos en cuanto a los métodos de enseñanza.

3.5 **Hipótesis**

Se busca probar que existe un público objetivo que desea calificarse profesionalmente en conocimientos administrativos, sistemas de gestión pública y contrataciones para prestación de servicio al Estado peruano con conocimientos altamente competitivos en el mercado laboral.

4 MARCO TEÓRICO

4.1 Metodología de ciencia de datos

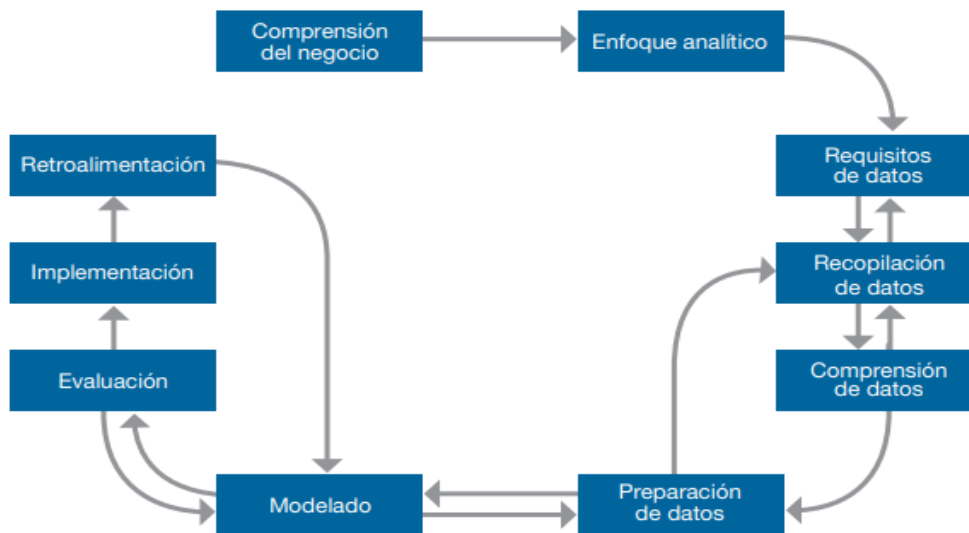


Figura 5. Metodología de datos

Fuente: IBM Analytics (s.f.)

Etapa 1: Comprensión del negocio

Todos los proyectos comienzan con la comprensión del negocio. Los promotores de negocios que necesitan la solución analítica desempeñan el papel más importante en esta etapa, al definir el problema, los objetivos del proyecto y los requisitos de la solución desde una perspectiva empresarial. Esta primera etapa sienta las bases para que el problema empresarial sea resuelto con éxito. (IBM Analytics, s.f., p. 3)

Etapa 2: Enfoque analítico

Cuando el problema empresarial se haya establecido claramente, el científico de datos podrá definir el enfoque analítico para resolver el problema. Esta etapa implica expresar el problema bajo el contexto de las técnicas estadísticas y de aprendizaje automático, para que la organización pueda identificar las más adecuadas para el resultado deseado. (IBM Analytics, s.f., p. 3)

Etapa 3: Requisitos de datos

“El enfoque analítico elegido determina los requisitos de datos. Más concretamente, los métodos analíticos a utilizar requieren de determinados contenidos de datos, formatos y representaciones, orientados por el conocimiento en el dominio” (IBM Analytics, s.f., p. 4).

Etapa 4: Recopilación de datos

En la etapa inicial de recopilación de datos, los científicos de datos identifican y reúnen los recursos de datos disponibles (estructurados, no estructurados y semiestructurados) y relevantes para el dominio del problema. Por lo general, deben elegir si realizan inversiones adicionales para obtener elementos informativos menos accesibles. (IBM Analytics, s.f., p. 4)

Etapa 5: Comprensión de datos

Después de la recopilación de datos inicial, los científicos de datos suelen utilizar estadísticas descriptivas y técnicas de visualización para comprender el contenido de los datos, evaluar su calidad y descubrir *insights* iniciales sobre ellos. Para llenar los huecos es posible que sea necesario volver a recopilar datos. (IBM Analytics, s.f., p. 4)

Etapa 6: Preparación de datos

Esta etapa abarca todas las actividades para construir el conjunto de datos que se utilizará en la subsiguiente etapa de modelado. Entre las actividades de preparación de datos están la limpieza de datos (tratar con valores no válidos o que faltan, eliminar duplicados y dar un formato adecuado), combinar datos de múltiples fuentes (archivos, tablas y plataformas) y transformar los datos en variables más útiles. (IBM Analytics, s.f., p. 4)

Etapa 7: Modelado

La etapa de modelado utiliza la primera versión del conjunto de datos preparado y se enfoca en desarrollar modelos predictivos o descriptivos según el enfoque analítico previamente definido. En los modelos predictivos, los científicos de datos utilizan un conjunto de capacitación (datos históricos en los que se conoce el resultado de interés) para construir el modelo. (IBM Analytics, s.f., p. 4)

Etapa 8: Evaluación

“Durante el desarrollo del modelo y antes de su implementación, el científico de datos evalúa el modelo para comprender su calidad y garantizar que aborda el problema empresarial de manera adecuada y completa” (IBM Analytics, s.f., p. 5).

Etapa 9: Implementación

Cuando el modelo satisfactorio ha sido desarrollado y aprobado por los promotores del negocio, se implementa en el entorno de producción o en un entorno de pruebas comparable. Por lo general, se implementa de forma limitada hasta que su rendimiento se haya evaluado completamente. Su implementación puede ser tan fácil como generar un informe con recomendaciones, o tan enrevesado como incrustar el modelo en un complejo proceso de puntuación y de flujo de trabajo administrado por una aplicación personalizada. (IBM Analytics, s.f., p. 5)

Etapa 10: Retroalimentación

Al recopilar los resultados del modelo implementado, la organización obtiene retroalimentación sobre el rendimiento del modelo y su impacto en el entorno en el que se implementó. Por ejemplo, la retroalimentación puede ser en forma de porcentajes de respuesta a una campaña promocional dirigida a un grupo de clientes que ha sido identificado por el modelo como respondedores de alto potencial. Los científicos de datos pueden analizar esta retroalimentación para ajustar el modelo para mejorar su precisión y utilidad. (IBM Analytics, s.f., p. 5)

Calidad de datos

Para poder adecuar este punto a la investigación en curso, primero se debe tener claro a qué se refiere la calidad de datos.

La calidad de los datos es más que una percepción. No existen estandarizaciones, ni una talla única en lo que se refiere a *data quality*. Mantener la exactitud y la integridad de todos los tipos de datos en toda la organización es trabajar por su aptitud para cumplir con su propósito en un contexto dado, implica garantizar que cada dato reúne todos los atributos necesarios. (PowerData, s.f., párr. 5)

✓ Exactitud

- Integridad.
- Actualización.
- Relevancia.
- Coherencia.
- Confiabilidad.
- Presentación apropiada.
- Accesibilidad.

En relación con lo mencionado acerca de la calidad de datos, se puede precisar que la data empleada en esta investigación se encuentra correctamente actualizada y limpia. Por tanto, no existe ningún registro duplicado que pueda causar algún tipo de error en el análisis cuantitativo de la investigación, dado que la base con la que se cuenta proviene de la misma organización (CCDG); en ese sentido, se la considera una fuente confiable. Además, dentro de una organización, la calidad de los datos es esencial para la consistencia del *reporting*, para la confianza de los usuarios y para la eficacia de los procesos operativos y transaccionales.

4.2 Gobierno de datos

La gestión de la calidad de los datos (Data Quality Management) es una forma de administración que abarca desde la definición y designación de roles hasta el

despliegue de funciones, de la definición de políticas y responsabilidades al establecimiento de procedimientos para la adquisición, mantenimiento, disposición y distribución de datos. (PowerData, s.f., párr. 12)

Un enfoque eficaz de la gestión de la calidad de los datos del Centro de Capacitación y Desarrollo Global, comprende tanto elementos reactivos, que incluyen la gestión de problemas en los datos situados en bases de datos existentes; como elementos proactivos. (PowerData, s.f., párr. 13)

Los anteriores son los que tienen que ver con esto:

- **Establecimiento de la gobernanza:** donde se estipula la normativa en la cual se basa la transferencia de data dentro de la organización.
- **Identificación de las funciones y responsabilidades:** en este punto, la empresa determina cuáles son los cargos que pueden tener autonomía sobre la data y qué roles pueden ejercer sobre esta.
- **Creación de las expectativas de calidad, así como de las estrategias empresariales de apoyo:** se requiere de mucha cultura de sensibilidad en cuanto al trato de los datos de los clientes dentro de la organización.
- **Implementación de una plataforma técnica que facilite estas prácticas empresariales:** el área específicamente encargada de esta distribución de datos. (PowerData, s.f., párr. 14)

Por eso, para que una iniciativa de gestión de la calidad de los datos tenga éxito dentro del Centro de Capacitación y Desarrollo Global, debe garantizarse la cooperación entre las áreas de IT y negocio. Esta asociación es importante porque, si bien los perfiles técnicos se encargarán de la construcción y el control del entorno, los usuarios de negocio serán los propietarios de los datos y, a partir de la aceptación ese rol, asumirán una responsabilidad con la organización y sus activos informacionales. (PowerData, s.f., párr. 15)

“Así, desde IT se llevarán a cabo todas las acciones necesarias para adquirir, mantener, difundir y poner a disposición de quien corresponda los activos de datos electrónicos de una organización” (PowerData, s.f., párr. 16).

4.3 Diccionario de datos

Este punto se trabajó con base en la data depurada; se consideró el medio de pago, los ingresos, el nombre de los cursos, la fecha de pago y el año en que este se realizó.

Tabla 3. Diccionario de datos

	Tipo	Longitud	Nulo	Índice	Único	Formato	Descripción	Regla de Negocio
ID	Cadena	5	No	Si	Si	ID###	Identificador	Ninguna
Medio de pago	Entero	50	NO	NO	NO		Banco donde se realizó el pago	validos BCP, INTERBANK, SCOTIABANK, BN, BBVB, OFICINA
Ingreso	Decimal	50	NO	NO	NO	Entero 10, decimal 2	pago realizado por cada estudiante	Debe ser positivo y mayor a 100
Nombre del curso	Cadena	150	NO	NO	NO		Nombre del curso inscrito	Es el nombre correcto que viene del Regulador. Con este se debe actualizar la base de la empresa
Fecha pago	Fecha	12	Si	NO	NO	DD/MM/AAAA	Fecha de pago del curso	validos año 2018 y 2019
Año	Fecha	5	NO	NO	NO	AAAA	Año que se realizó el pago	validos año 2018 y 2019
Programa	texto	50	NO	NO	NO		Programa y/o servicio brindado	Ninguna
Modalidad	texto	30	NO	NO	NO		Modalidad de enseñanza	Presencial o virtual
Vendedor	texto	30	NO	NO	NO		Asesor que captó y cerró la venta	Ninguna
Medio de contacto	texto	30	NO	NO	NO		Medio por el cual se contactó al alumno	validos solo : Whatsapp, oficina , llamada, facebook. Correo, zopim, web, recomendado
Edad	Entero	5	NO	NO	NO	Entero 10, decimal 2	Edad del alumno	Igual o mayor a 18 años
Clase	texto	5	NO	NO	NO		Clase social	Solo A- B -C

Fuente: elaboración propia

5 ARQUITECTURA DE DATOS

5.1 Arquitectura de datos funcional



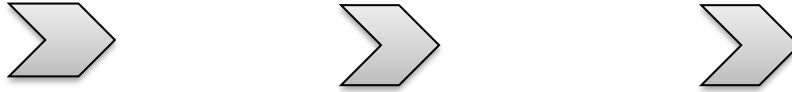


Figura 6. Arquitectura de datos funcional

Fuente: elaboración propia

5.2 Arquitectura de datos tecnológica



Figura 7. Arquitectura de datos tecnológica

Fuente: elaboración propia

a) Insumos

La data utilizada para hacer el análisis de los datos proviene de las siguientes plataformas:

- **ERP:** el Sistema Integrado de Gestión (SIG) que utilizan en la empresa es el SAP, y se tomó la data transaccional del área de finanzas con respecto a las ventas concretadas en los años 2018 y 2019.
- **Data de medios sociales:** se extraen los datos disponibles por medios sociales, n como WhatsApp, Facebook e Instagram, donde los clientes hacen sus registros y las ventas son concretadas.
- **Data de página web:** la empresa cuenta con una aplicación dentro de la página web www.desarrolloglobal.pe, donde los clientes pueden hacer sus consultas y recibir asesoría personalizada por medio de Zopim. Esta es otra de las formas de recolectar la información de potenciales usuarios.

b) Procesamiento

La integración de datos se implementa con IBM InfoSphere, que cuenta con un *data warehouse*; este es un *software* que aloja los datos internos y externos. Estos son extraídos, recuperados y presentados de forma unificada de las áreas de marketing, ventas, operaciones y finanzas. Esta herramienta ayuda a la toma de decisiones desde cualquier área de la empresa, con lo que se obtiene una información integrada y global y se encuentran relaciones e implantación de sistemas de gestión integral en relación con los clientes de la empresa.

c) Almacenamiento

Los datos son almacenados en Power BI, para que los colaboradores responsables puedan tener acceso total al sistema por medio del *desktop* o la aplicación en el celular. Esta herramienta permite recolectar todas las bases de información de las distintas áreas, conectarse con las demás fuentes diversas de información, e incluso compartirlas.

d) Destino

Estos datos deben usarse para mejorar la toma de decisiones, tener una mejor planificación empresarial y generar reportes de ventas consolidadas, predicción de análisis y modelos de datos que ayuden a cumplir los objetivos de la empresa usando la herramienta Power BI.

e) Tecnología, herramientas e infraestructura

Se puede utilizar Power BI para la visualización de datos, porque cuenta con una versión gratuita en el *desktop* y es una herramienta amigable para el usuario. Además, es una solución especializada en inteligencia empresarial que permite unir diferentes fuentes de datos y analizar estos para su uso; y presentarlos a través de paneles e informes, de manera que puedan ser consultados de una manera muy fácil, atractiva e intuitiva. Esta explotación de datos a través de paneles e informes implica la posibilidad de que sean compartidos por muchos usuarios de una misma empresa u organización.

Ventajas de usar Power BI:



Figura 8. Power BI

Fuente: (Oyelowo, 2018)

- Plataforma especializada en visualizaciones.
- Herramienta de inteligencia de negocio y análisis.
- Solución sólida y comprobada de BI que tiene una gran base de usuarios, un número que sigue en aumento.
- Un potente servicio y con todas las funciones basadas en la nube que ayuda a los usuarios del negocio.

5.3 Comprensión y preparación de los datos

5.3.1 Revisión y depuración de datos que se usaron en el estudio

Los datos que se utilizaron son confidenciales, puesto que son internos de la organización; por ello, se debe omitir cierta información delicada del usuario, como su nombre y el DNI

(documento nacional de identidad), del cual solo se extrajo la fecha de nacimiento para calcular la edad del estudiante. Así, de una base completa de los datos de alumnos, se procedió con la limpieza y depuración de muchos de los puntos que se estipulaban en la data, pues solo se requerían los siguientes:

- ID del cliente.
- Fecha de pago.
- Año de inscripción.
- Medio de pago.
- Ejecutivo de venta.
- Programa inscrito.
- Monto de pago.
- Edad.
- Clase económica.
- Medio de contacto.
- Ventas del 2018 y 2019.

Una vez separada y seleccionada esta data, se estructuraron los gráficos en cuanto a cada tema y punto a tratar, con el fin de responder las preguntas de Data Science y llegar a una conclusión con respecto al estudio en el presente trabajo de investigación.

5.4 Requisitos y recopilación de datos

5.4.1 Determinación de datos necesarios para responder a las preguntas

Los datos que requeridos para responder las preguntas planteadas son los siguientes:

- Ventas del 2018 y 2019.
- ID del cliente.
- Fecha de pago.

- Año de inscripción.
- Medio de pago.
- Ejecutivo de venta.
- Programa inscrito.
- Monto de pago.
- Edad.
- Clase económica.
- Medio de contacto.

5.5 Ubicación de fuentes de datos

La empresa cuenta con un sistema integrado que se encarga de la recopilación de la información sobre las ventas que ingresa el área comercial. Se hizo la extracción de la data necesaria (correspondiente a los años 2018 y 2019) para desarrollar este proyecto, con la finalidad de cumplir los objetivos trazados.

- Información sobre el medio utilizado para contactar a los alumnos matriculados en los años 2018 y 2019.
- Datos de los programas que tuvieron mayor demanda en los años 2018 y 2019.
- Datos sobre la modalidad, virtual o presencial, que seleccionaron los alumnos al momento de la matrícula en los dos últimos años.

Tabla 4. Fuentes de datos

Datos	fuer	ubicación	Cuento con	Plan de
	te		acceso	acción
✓ ID del cliente	Fuente interna	Sistema integrado de la empresa	Sí	
✓ Fecha de pago				

-
- ✓ Año de inscripción
 - ✓ Medio de pago
 - ✓ Ejecutivo de venta
 - ✓ Programa inscrito
 - ✓ Monto de pago
 - ✓ Edad
 - ✓ Clase económica
 - ✓ Medio de contacto
-

Fuente: elaboración propia

5.6 Presentación de datos recopilados

Volumen diario de nuevos procesos de contratación que son publicados en el Sistema Electrónico de Contrataciones del Estado (Seace)

La cantidad de licitaciones publicadas diariamente es similar entre 2017 y 2018, y también se nota el incremento mensual que va cuesta arriba hacia el final del año.

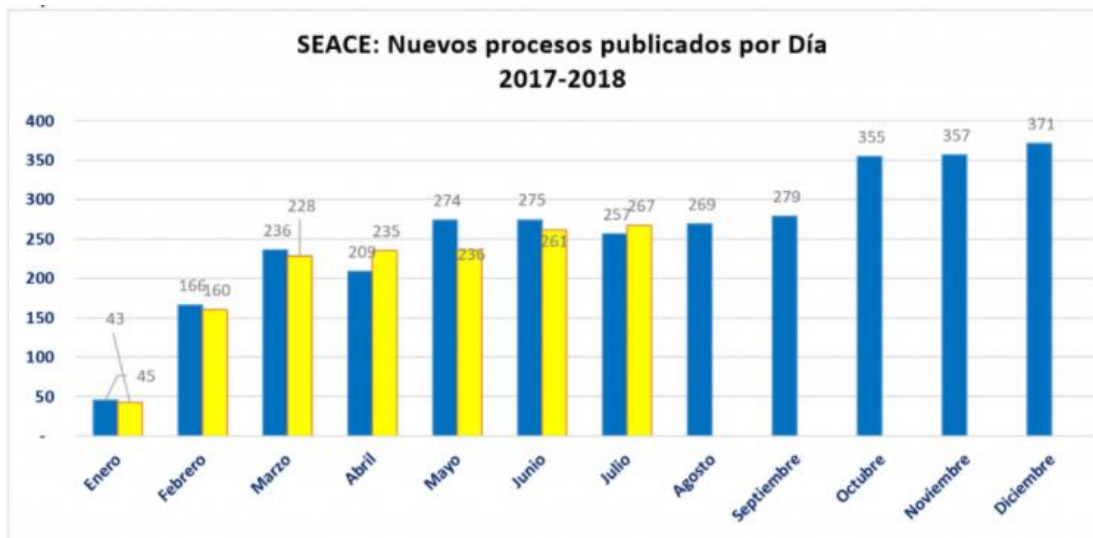


Figura 9. Seace: nuevos procesos publicados por día

Fuente: (LumenSoft, 2018)

Cantidad de procesos de contratación publicados mensualmente

Como se puede apreciar, el volumen de los procesos en contratación va en aumento conforme se acerca el cierre del año.

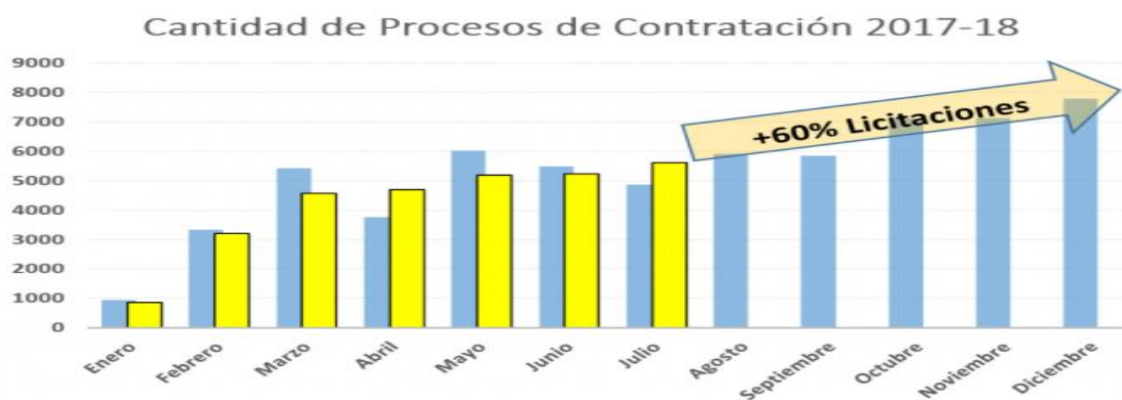


Figura 10. Contrataciones 2017-2018

Fuente: (LumenSoft, 2018)



Figura 11. Estado de procesos de contratación 2018

Fuente: elaboración propia

5.7 Uso de herramientas para determinar posibles relaciones entre los datos

5.7.1 Determinando el punto de partida

Para determinar si existe relación entre los datos, el equipo utilizó la herramienta de Excel, donde, con apoyo del área administrativa del CCDG, quienes facilitaron la data del alumno y de los cursos, se pudieron consignar los datos de la cantidad de alumnos matriculados, los cursos con mayor demanda y la cantidad de licitaciones por año. Con base en estos datos, se pudo determinar que sí existe dicha relación entre los datos, pues al darse un crecimiento en la cantidad de vacantes por licitaciones, hay una mayor posibilidad de que los alumnos

ejerzan su especialidad una vez que egresen del CCDG. Asimismo, se pudo describir, según la tendencia, qué tipo de especialización es la que tiene mayor réplica en el ámbito laboral.

5.7.2 Determinación de las herramientas de visualización a utilizar para el proyecto

La herramienta utilizada para determinar las posibles relaciones entre los datos fue Power BI, puesto que es una plataforma amigable con el usuario y permite subir diferentes formatos de archivos y realizar presentaciones atractivas para la creación de *dashboards* empresariales, a fin de sacar conclusiones. Además, es de fácil acceso, tiene una inscripción gratuita y gran contenido instructivo de la plataforma en la web.



Figura 12. Power BI

Fuente: (Eucariota, s.f.)

6 PRODUCCIÓN Y ANÁLISIS EXPLORATORIO DE DATOS

6.1 Plan de muestreo y tamaño de la muestra

6.1.1 Definir la población objetivo

Para el plan de muestreo se seleccionaron los alumnos que estudiaron en el CCDG. A partir de estos, se tomó la información requerida de las características relevantes que ayudaron a resolver el problema de investigación. Los alumnos fueron hombres y mujeres, mayores de

18 y menores de 55 años principalmente, de niveles socioeconómicos A, B y C, que viven en la ciudad de Lima.

6.1.2 Identificar las unidades de prueba

Los alumnos son hombres y mujeres, mayores de 18 y menores de 55 años principalmente, de los niveles socioeconómicos A, B y C, que viven en la ciudad de Lima.

6.1.3 Seleccionar la técnica de muestreo

Se usó el muestreo probabilístico, donde cada elemento de la población tiene la misma probabilidad de ser seleccionado para participar en el estudio. Si bien es cierto que este tipo de muestreo favorece el análisis estadístico porque permite calcular el margen de error y el intervalo de confianza, también representa con mayor precisión a la población de los estudiantes de los años 2018 y 2019, por lo que se considera una muestra heterogénea.

6.2 Tamaño de la muestra

Para el tamaño de la muestra se seleccionó a un subgrupo de los alumnos de la población seleccionada para participar en el estudio; estos corresponden a quienes estudiaron en el CCDG en los años 2018 y 2019 y participaron en los cursos, diplomas y diplomados.

Calculadora de muestra

Nivel de Confianza 95% 99%

Margen de Error:

Población:

Tamaño de Muestra:

Figura 13. Cálculo de la muestra

Fuente: (QuestionPro, s.f.)

6.3 Diseño del estudio de muestreo

Se utilizó el muestreo probabilístico aleatorio simple, esto quiere decir que los individuos pueden ser seleccionados más de una vez en la muestra; ello, debido a que se implementó un diseño de estudio de muestreo con reposición, pues hubo alumnos que se matricularon en más de un curso. Así, el hecho de que se seleccione a un usuario al azar para la muestra no evita que ese mismo estudiante puede volver a ser seleccionado; de esta manera, se asegura la obtención de muestras representativas, donde la única fuente de error que afecte a los resultados es el azar.

6.4 Ejecución del muestreo

Para llevar a cabo el muestreo, se tomó en consideración una población de 5890 estudiantes extraída de la base de datos de la empresa, quienes corresponden a los dos últimos años. A fin de hallar el tamaño de la muestra, se consideró un nivel de precisión de 5 y un nivel de

confianza del 95 %; a Z se le asignó un valor de 1,96 y una desviación estándar de 60. El resultado se muestra a continuación.

- Nivel de precisión: $D = 5$
- Nivel de confianza: $CL = 95 \%$
- Valor Z: 1,96
- Desviación estándar de la población: $\sigma = 60$

Consecuentemente, se calculó el tamaño de la muestra, utilizando la fórmula del error estándar:

$$n = (\sigma^2 z^2) / D^2 \quad n = (60^2 \cdot 1,96^2) / 5^2$$

$$n = 545$$

6.5 Resumen y examen de la distribución de variables

Para determinar la distribución de la variable adecuada para este tipo de datos, se utilizó un modelo teórico con una distribución normal capaz de aproximar el valor de una variable aleatoria continua a una situación de mayor demanda de cursos, diplomas y diplomados en la institución.

También se consideraron todos los datos de variables. En total, se contó con 11 de estas, entre las cuales se tuvieron las cualitativas: el nivel socioeconómico de los estudiantes (variable cualitativa ordinal); la modalidad de enseñanza preferida (variable cualitativa nominal); y el programa, el vendedor, el medio de contacto y el mes de ingreso. Igualmente, las variables cuantitativas: el ingreso por alumno y el ingreso total por año y edad (variables cuantitativas continuas). Referente a los estudiantes en los años 2018 y 2019, que suman 5890, se identificaron las preferencias de los estudiantes para los cursos a futuro.

6.6 Distribución de la muestra: interpretación del comportamiento

Esta muestra se escogió con base en los dos últimos años, con un total de 5890 estudiantes del total de cursos dictados en el CCDG. Según el cálculo, la muestra resultó en 545 alumnos del subgrupo trabajado.

El objetivo es efectuar una generalización de los resultados de la muestra a la población. Inferir o adivinar el comportamiento de nuestra población a partir del conocimiento de una muestra. Para ello es necesario conocer las distribuciones de probabilidad de ciertas funciones de las muestras que constituyen variables aleatorias asociadas al experimento aleatorio, selección de una muestra al azar de una población. Estas variables aleatorias denominadas estadísticos muestrales, porque se basan en el comportamiento de las muestras, asignan a cada muestra del espacio muestral, constituido por todas las muestras posibles, un número real que es un resumen estadístico de la muestra. (UGR, s.f., p. 1)

6.7 Presentación de datos a utilizar para el estudio

Para el CCDG, se seleccionaron los datos recopilados de la empresa con la ayuda de la página web y el área de administración del mismo centro, con los datos de los dos últimos años (2019 y 2018). Estos se consiguieron a través de una base de datos amplia, de la cual se tenía información no relevante, como la información personal del alumno. A partir de ahí, se procedió a ordenar de manera selectiva la data.

Se procedió con la depuración y limpieza de los datos de forma ordenada, teniendo en cuenta puntos importantes y de gran utilidad para el análisis. Dentro de la información se tuvo el N° de ID del cliente, la fecha de pago, el año de inscripción, el medio de pago, el programa inscrito, el monto de pago, la edad, la clase económica y el medio de contacto.

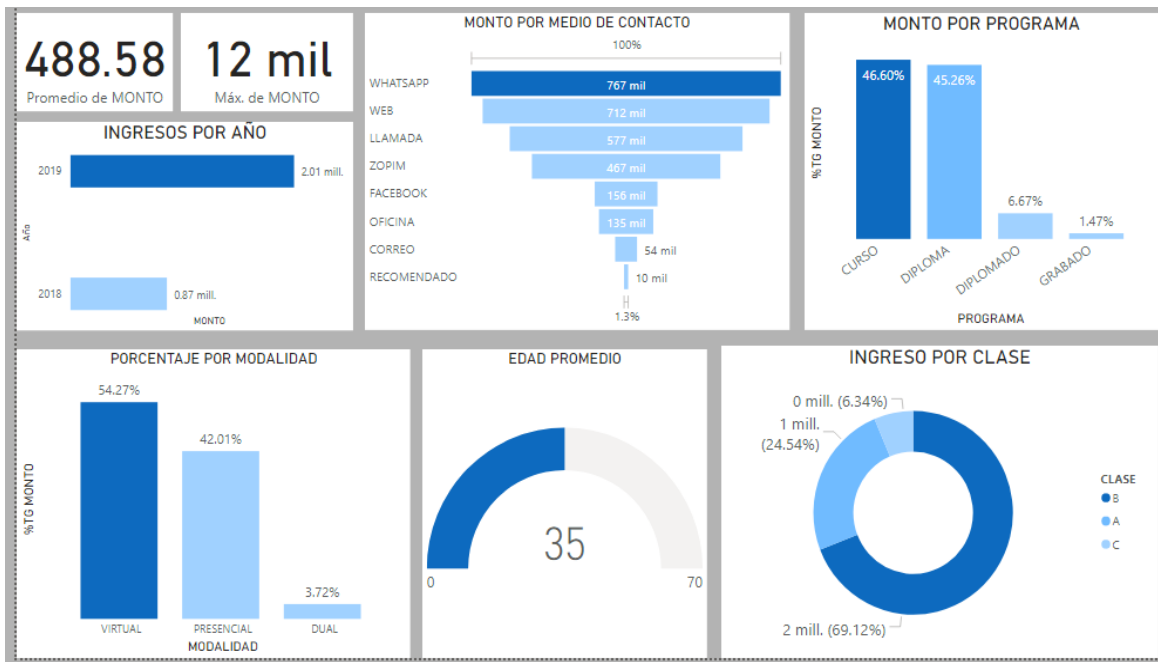


Figura 14. *Dashboard* de desarrollo global

Fuente: elaboración propia

En el siguiente gráfico, se puede ver el ingreso total por años; se observa que el del año 2019 es tres veces mayor que el del 2018, lo que indica que hay más interés del público objetivo para llevar cursos y diplomados. El siguiente gráfico muestra el ingreso según la clase de los estudiantes, y se observa que la clase B es la que tiene mayor participación, con casi un 70 %, seguido por la clase A, con el 24,5 %.

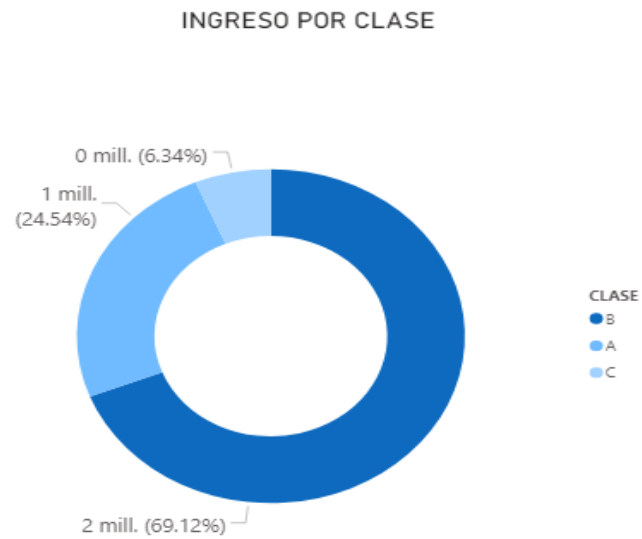


Figura 15. Ingreso por clase

Fuente: elaboración propia

También se saca la edad promedio de los estudiantes para conocer la tendencia de los matriculados, y se evidencia que es de 35 años.

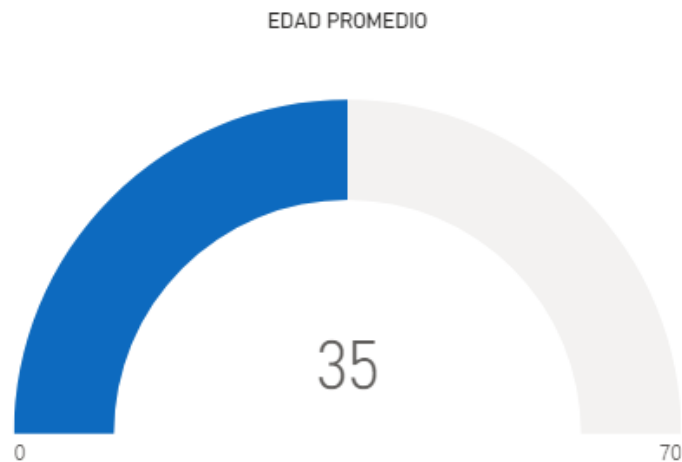


Figura 16. Edad promedio de estudiantes

Fuente: elaboración propia

En esta visualización, se observan los ingresos de acuerdo con la opción de modalidad. Como resultado, se tiene que la más elegida ha sido la modalidad virtual.

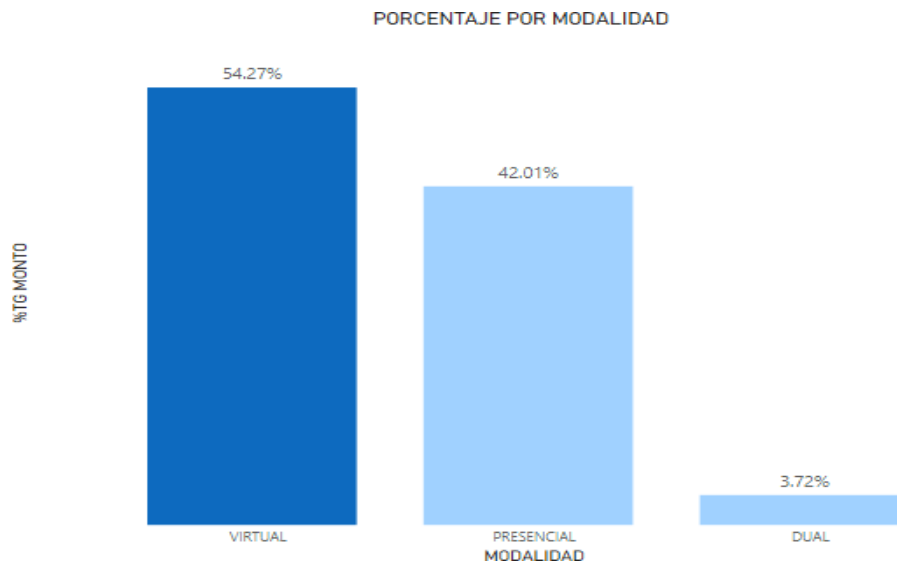


Figura 17. Porcentaje por modalidad de estudio

Fuente: elaboración propia

En este gráfico de barras, se visualiza que los mayores ingresos obtenidos han sido por curso y por diploma; y en un porcentaje muy inferior, por diplomado.

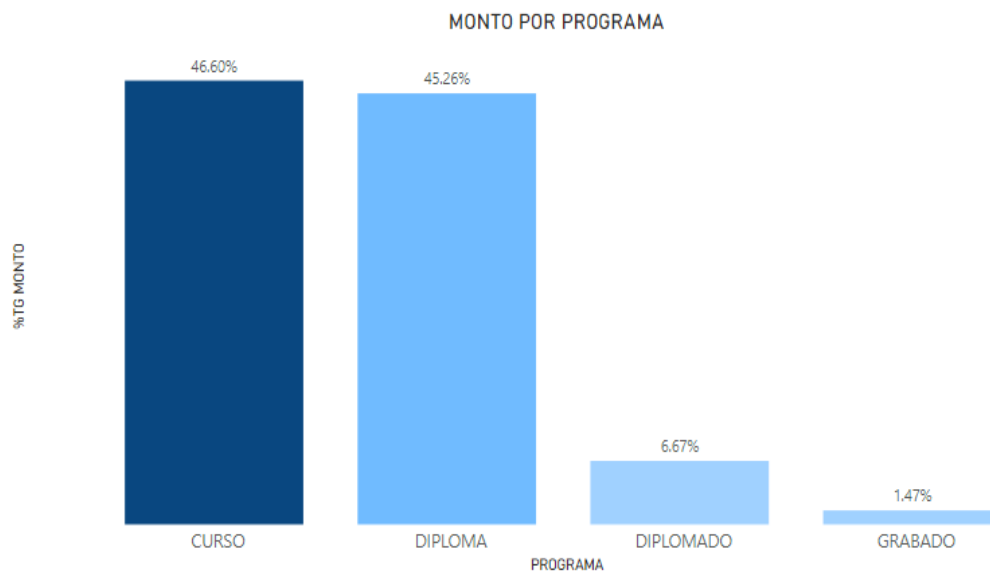


Figura 18. Monto por programa

Fuente: elaboración propia

En el siguiente gráfico, se muestra el ingreso según el medio de contacto, con lo que se nota que la tendencia para cerrar una venta efectiva es por medio de WhatsApp (767 000 soles). Esta opción es superior frente a las demás.

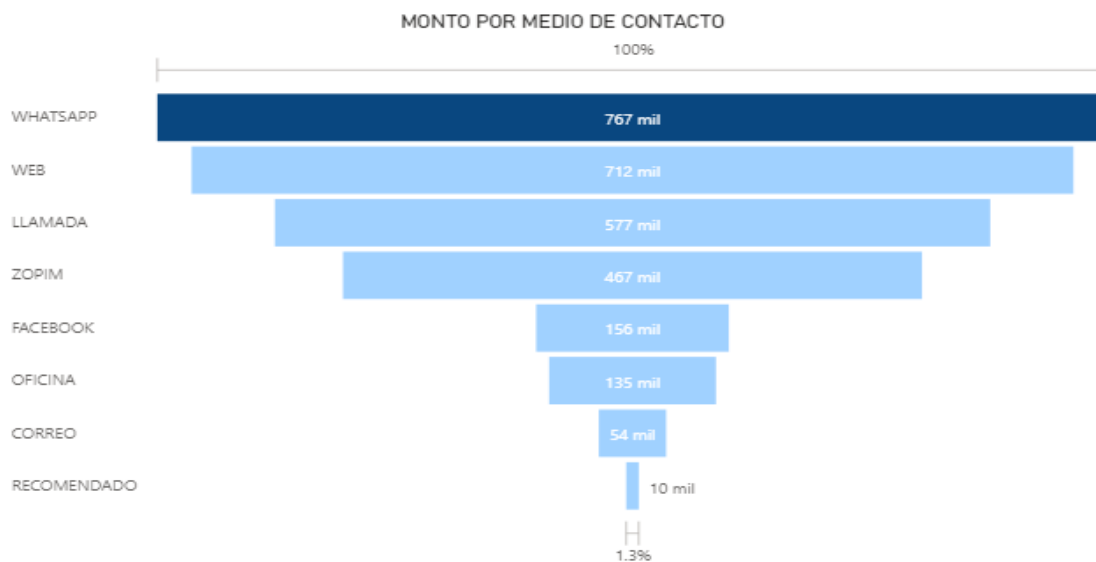


Figura 19. Monto por medio de contacto

Fuente: elaboración propia

En este gráfico, se observa que la ejecutiva con mayores ventas fue Jelissa, con un monto total de 7000 soles. La siguiente vendedora con más ventas fue Andrea.

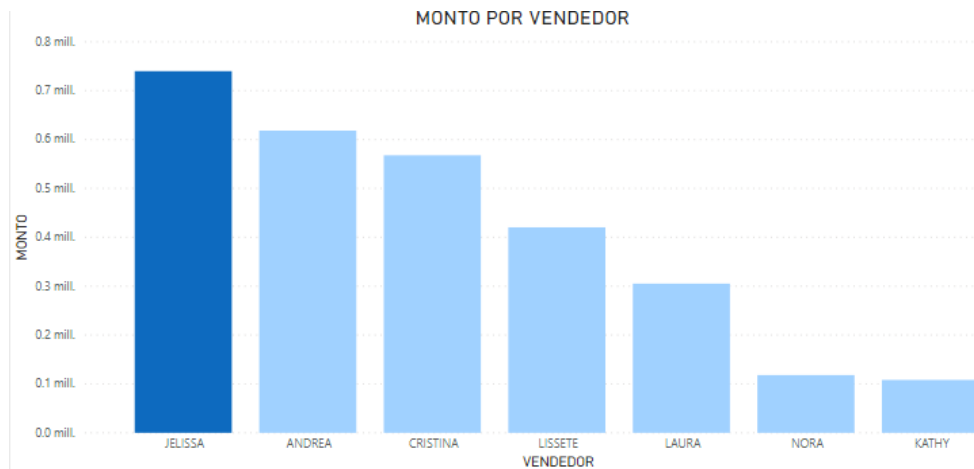


Figura 20. Monto por vendedor

Fuente: elaboración propia

6.8 Resultado del análisis de relaciones

Para este punto, se relacionaron las variables de la modalidad de enseñanza, estas son, presencial y virtual. Lo que se busca es encontrar un vínculo entre los programas dictados – cursos, diplomas y diplomados– en el CCDG. En los siguientes gráficos, se visualiza la relación entre la variable de programas de enseñanza y la probabilidad de elección del alumno.



Figura 21. Relación de variables de modalidad virtual y programa

Fuente: elaboración propia

Cuando el tipo de curso es diploma, es 1,38 veces más probable que la modalidad de enseñanza sea virtual, en comparación con el resto de los valores de programa. Este elemento influyente contiene aproximadamente el 35,06 de los datos.

Cuando la variable de programa es diplomado, es 1,34 veces más probable que la modalidad elegida por el estudiante también sea virtual, en comparación con el resto de las variables.



Figura 22. Relación de variables modalidad presencial con programa

Fuente: elaboración propia

Cuando la variable de programa es curso, es 1,64 veces más probable que la modalidad sea presencial.

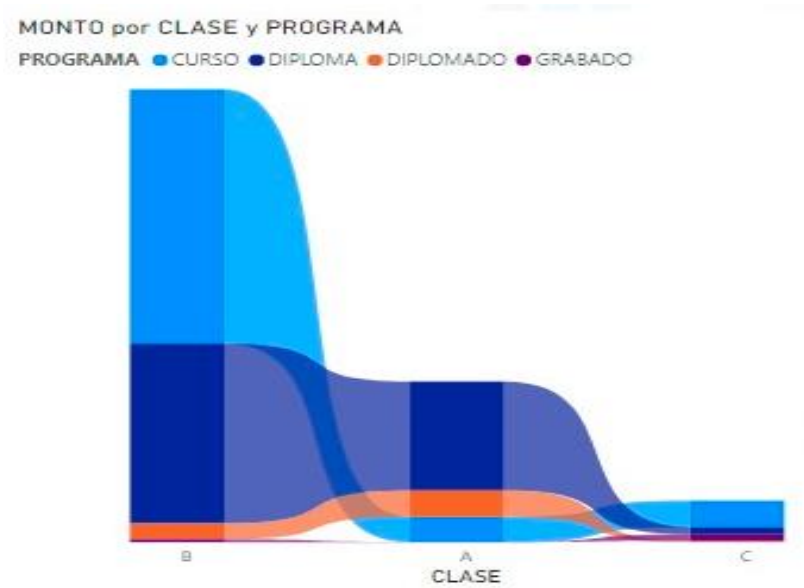


Figura 23. Relación entre la variable clase y programa

Fuente: elaboración propia

En este gráfico, se relacionan las clases A, B y C con los programas detallados en el Anexo 1. Se puede visualizar e interpretar que la clase B tiene mayor preferencia por los cursos que se dictan en el centro.

6.9 Estudio probabilístico de los datos obtenidos

Se decidió, como equipo, basarse en el siguiente tipo de muestreo para obtener los datos:

6.9.1 Técnica de muestreo probabilístico

Al hablar de muestreo probabilístico, siempre se cumplen dos condiciones:

1. “Todos los elementos de la población tienen una probabilidad mayor a cero de ser seleccionados en la muestra.
2. La probabilidad de inclusión de cada elemento en la muestra se conoce de forma precisa” (Ochoa, 2015, párr. 6).

Tabla 5. Datos con estadística

MEDIA	495.14	PROMEDIO DE LOS DATOS
MEDIANA	404.50	VALOR CENTRAL DE LOS DATOS
MODA	397.00	DATO QUE MÁS SE REPITE
DESVIACIÓN ESTANDAR	309.36	MEDIDA DISPERSA MAS COMÚN
RANGO	100.00	VALOR MÍNIMO
	3191.00	VALOR MÁXIMO

Fuente: elaboración propia

La herramienta utilizada para los cálculos del cuadro anterior fue Excel, debido a la cantidad de datos con los que se contaba. Estos fueron extraídos de las variables del monto de pago que los estudiantes hacían a las cuentas de la empresa por los diversos canales habilitados, a fin de asistir al curso o diploma elegido. Con respecto al hallazgo de la media y el valor medio, el cálculo se realizó con el promedio de los números; en este caso, se sumaron los números y se dividieron por la cantidad de estos. Según la base de datos, se tenían 5985, lo que daba como resultado 495,14. En relación con la mediana, se ordenaron los números según su valor, y se identificó el del medio, este fue 404,50.

Asimismo, se procedió a hallar la moda, esta es simplemente el valor que aparece más veces. Para calcularla, se ordenaron los números; ello dio como resultado 397, correspondiente a

los pagos de los cursos más demandados, como el de Sistema Integrado de Gestión Administrativa (SIGA) y el de Seace.

Según los cálculos, la desviación estándar es igual a 309,36; esta es la medida más común, pues indica qué tan dispersos están los datos con respecto a la media. Finalmente, se tuvieron los rangos con un dato mínimo de 100 y máximo de 3191; y si estos se restan, se tiene un resultado de 3091. Esta es una medida de dispersión, dado que muestra cómo los datos individuales pueden diferir de la media.

FECHA	AÑO	MEDIO DE PAGO	MONTO	PROGRAMA	MODALIDAD	VENDEDOR	MEDIO DE CONTACTO
2/01/2018	2018	PAYU	257.00	CURSO PREPARATORIO OSCE II	VIRTUAL	DANIELA	ZOPIM
2/01/2018	2018	PAYU	457.00	CURSO PREPARATORIO OSCE	PRESENCIAL	JELISSA	WHATSAPP
2/01/2018	2018	BCP	457.00	CURSO SIAF BASICO	PRESENCIAL	JELISSA	WHATSAPP
2/01/2018	2018	BN	257.00	CURSO PREPARATORIO OSCE II	VIRTUAL	MARISOL	FACEBOOK
2/01/2018	2018	BCP	257.00	CURSO OSCE II	VIRTUAL	MARISOL	WEB
2/01/2018	2018	INTERBANK	257.00	CURSO OSCE II	VIRTUAL	MARISOL	WEB
3/01/2018	2018	BN	257.00	CURSO PREPARATORIO OSCE II	VIRTUAL	DANIELA	ZOPIM
3/01/2018	2018	BN	248.50	CURSO OSCE II	VIRTUAL	DANIELA	ZOPIM
3/01/2018	2018	BCP	397.00	CERTIFICADO INVIERTE PERÚ	VIRTUAL	DANIELA	FACEBOOK
3/01/2018	2018	BN	141.50	CURSO PREPARATORIO OSCE II	VIRTUAL	MARISOL	WEB
3/01/2018	2018	PAYU	257.00	CURSO PREPARATORIO OSCE II	VIRTUAL	MARISOL	WEB
3/01/2018	2018	VISA	457.00	CURSO PREPARATORIO OSCE II	PRESENCIAL	JELISSA	LLAMADA
3/01/2018	2018	BCP	257.00	CURSO PREPARATORIO OSCE II	VIRTUAL	MARISOL	LLAMADA
4/01/2018	2018	BCP	397.00	CERTIFICADO INVIERTE PERÚ	VIRTUAL	DANIELA	WEB
4/01/2018	2018	BN	1297.00	DIPLOMADO GESTIO PUBLICA	VIRTUAL	MARISOL	WEB

Figura 24. Datos para análisis

Fuente: elaboración propia

La mediana es mayor que la moda, pues da como resultado 404,50, mientras que el dato de la moda es 397, es decir, el valor que más se repite. La mediana representa el valor de la variable de posición central en un conjunto de datos ordenados, y la moda, al ser menor, es el monto de pago con mayor frecuencia en la distribución de los datos. Los rangos están distribuidos de forma ascendente, desde el menor pago al mayor, con lo que se tienen los siguientes datos:

- Mínimo: 100
- Máximo: 3191

Se obtuvo como rango 3090, el cual representa la diferencia entre el valor máximo y el valor mínimo de un conjunto de datos. De acuerdo con este análisis, el conjunto de datos usado para hacer los cálculos había tenido una limpieza inicial meses atrás, debido a que estos se habían usado para hacer algunos gráficos pedidos en el curso; por ello, se tuvieron todos los datos completos y ordenados, con las celdas llenas. Esto ayudó a conseguir la información completa y hacer un mejor cálculo de las variables, a fin de contar con un mejor análisis.

7 TÉCNICAS DE CIENCIA DE DATOS

La técnica de ciencia de datos adaptada al presente proyecto es al siguiente:

Aprendizaje supervisado: se cuenta con datos históricos que generan una investigación predictiva en cuanto al conjunto de datos. Los algoritmos trabajan con datos etiquetados para encontrar una función con las variables de entrada y una etiqueta de salida adecuada, es decir, para predecir el valor de posibles alumnos matriculados en los cursos y diplomados que ofrece la empresa. Este tipo de aprendizaje utiliza dos técnicas, la de clasificación y la de

regresión. Se aplicó la técnica de regresión lineal para predecir la cantidad de alumnos que podrían inscribirse en los diversos programas y cursos del centro de capacitación.

Regresión lineal

Se diseñó una figura de regresión lineal comparando los meses desde enero de 2018 a diciembre de 2019; así, se utilizaron los montos totales por mes y se obtuvo el siguiente gráfico, el cual muestra una regresión positiva. Este gráfico permite establecer una fórmula específica de regresión:

$$y = bx + a$$

Con esta, al reemplazar x con el número del mes, se puede predecir de cuánto es el monto estimado que se puede vender. La fórmula obtenida es:

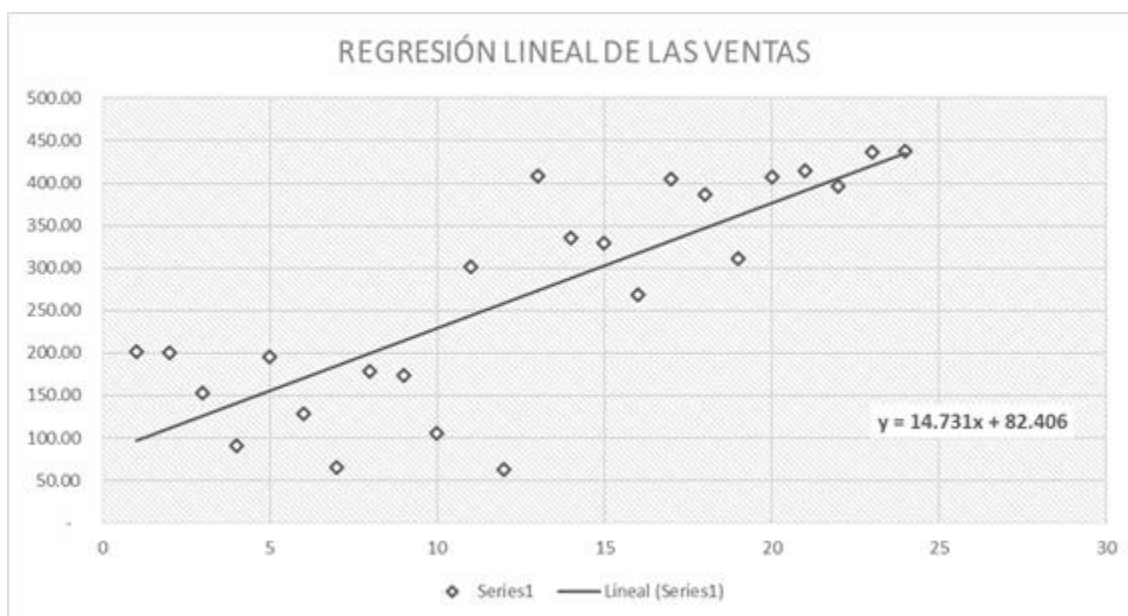


Figura 23. Regresión lineal

Fuente: elaboración propia

Tabla 6. Proyección 2020

Proyección 2020		
ID	Mes	N° de alumnos matriculados
25	Ene-20	451
26	Feb-20	465
27	Mar-20	480
28	Abr-20	495
29	May-20	510
30	Jun-20	524
31	Jul-20	539
32	Ago-20	554
33	Set-20	569
34	Oct-20	583
35	Nov-20	598
36	Dic-20	613
Total		6,380

13	Ene-19	409
14	Feb-19	335
15	Mar-19	329
16	Abr-19	268
17	May-19	405
18	Jun-19	387
19	Jul-19	311
20	Ago-19	407
21	Set-19	415
22	Oct-19	397
23	Nov-19	437
24	Dic-19	438

Histórico		
ID	Mes	N° de alumnos matriculados
1	Ene-18	202
2	Feb-18	201
3	Mar-18	153
4	Abr-18	91
5	May-18	195
6	Jun-18	128
7	Jul-18	65
8	Ago-18	179
9	Set-18	174
10	Oct-18	106
11	Nov-18	302
12	Dic-18	63

Fuente: elaboración propia

Con los datos mostrados en la tabla, se calculó la regresión; esto, tomando los registros de la data de los años 2018 y 2019 y el proyectado del año 2020. Los datos muestran el número de alumnos matriculados por mes.

7.1 El modelo y la historia

7.1.1 Determinación del modelo más apropiado para la descripción o predicción.

El modelo más apropiado es el predictivo; se sentaron las bases en los datos históricos para predecir el comportamiento del futuro. En tal caso, ello permitiría predecir la cantidad de alumnos a matricularse en los programas de capacitación para el año 2020; de esta manera, se podría planificar mejor la cantidad de cupos que se requieren para la apertura de cada uno de los cursos ofrecidos.

La demanda y venta de los cursos del CCDG están vinculadas con el interés y la necesidad del alumno o cliente. Para el presente trabajo, se ha establecido un enfoque en los factores principales que determinan la venta de cada uno de los cursos y programas de especialización; estos son los siguientes:

- **Planes y programas de capacitación**

El cliente evalúa qué tipo de plan y programa le conviene, de acuerdo con lo que desea cursar o especializarse. Además, este factor también depende de qué tan demandado sea el curso afuera, en el mercado laboral.

- **Presupuesto**

Este aspecto se enfoca en la economía y el alcance del cliente; este debe evaluar los costos del catálogo de cursos y especializaciones.

- **Metodologías didácticas**

Este factor va de la mano con la modalidad de enseñanza, la cual puede ser de tipo virtual o presencial; así, las metodologías que se aplican son diferentes para cada una.

- **Sistema automatizado de consultas**

Un factor relevante es el personal que se encarga de establecer el primer contacto con el cliente; este punto tiene mucho que ver con la rapidez y eficacia de la respuesta, debido a que, a menor margen de tiempo para la atención del cliente, mayor es la posibilidad de que este adquiera un curso, puesto que el tiempo es un recurso importante.

- **Tipo de especialización**

Con respecto a la especialización, esta se enfoca, mayormente, en el campo en el que se desempeña el cliente y posible alumno dentro del ámbito laboral; sin embargo, podrían darse casos en que este desee explorar o especializarse en un área diferente.

- **Formación de instructores**

La capacitación y el nivel de formación de los instructores o profesores del CCDG es un punto relevante, por no decir obligatorio, puesto que para el centro es muy importante brindar una educación de calidad.

- **Plan integral de formación**

Este último factor apunta al sistema general de educación que proyecta el CCDG; esto es muy importante para el cliente, porque engloba todos los puntos anteriores, su método de planificación y cómo se minimizan los riesgos.

8 DESCRIPCIÓN DE LOS HALLAZGOS

Con la información del gráfico consignado líneas abajo, se ha descubierto que los factores más relevantes que determinan si los posibles alumnos se matriculan en los cursos, programas y diplomados del CCDG son cuatro:

- El nivel de formación de la plana docente, que se refiere al nivel de capacitación y experiencia laboral de los profesores.
- La metodología utilizada, preferiblemente didáctica.
- El tipo de especialización y certificación para obtener una remuneración más elevada.
- Nivel de rapidez de las consultas por los diferentes canales de atención del CCDG.

8.1 Visualización, implementación y ajustes

8.1.1 Definición y justificación de los criterios a tomar para la visualización de los resultados

Para el presente trabajo, los factores considerados como más relevantes para construir las visualizaciones, las cuales podrían ayudar a responder las preguntas de investigación en relación con los objetivos planteados, son los siguientes:

- Elegir una herramienta para la visualización que sea amigable para el usuario y permita mostrar los hallazgos más resaltantes en función de lo que se quiere lograr en el proyecto.
- La plataforma debe tener bajo costo de utilización para su fácil acceso y análisis profundo de los datos y patrones.
- Las visualizaciones deben captar la atención y el interés del público espectador, con el fin mostrar y explicar las tendencias de los gráficos.

- Debe extraerse información importante para una amplia gama de escenarios de estudio; en este caso, para realizar una investigación predictiva y conseguir una mejor toma de decisiones.

Con base en la data, se construyeron las visualizaciones en la plataforma Power BI, debido a que había una familiaridad con esta herramienta; además, se consideró que esta cumplía con todos los atributos necesarios para adaptarla a las presentes variables, a fin de responder las preguntas de Data Science. Asimismo, se pensó en la consecución de los objetivos generales y específicos, con el propósito de llegar a conclusiones que permitieran tener mejores resultados para la organización.

8.2 Visualización de los hallazgos

Como se mencionó en el punto anterior, se utilizó Power BI para presentar las visualizaciones a modo de *dashboard*. Para ello, se construyeron los gráficos de manera independiente y, posteriormente, se consignaron las conclusiones:

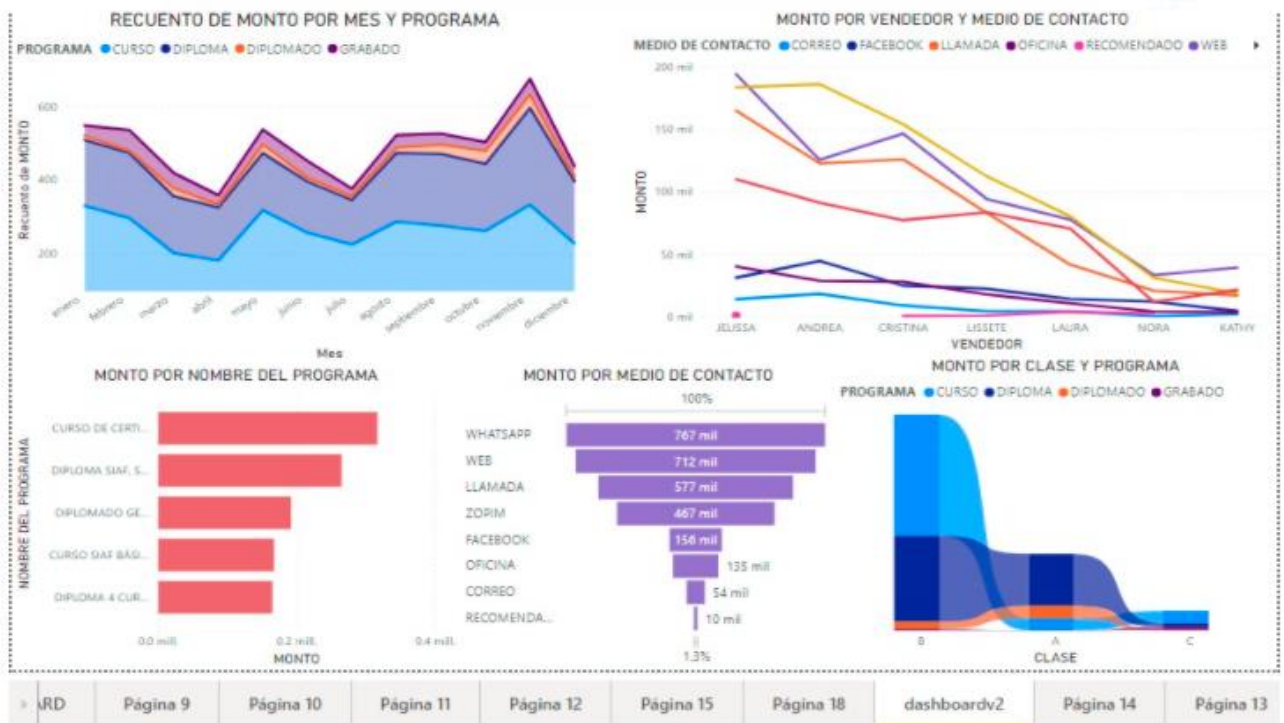


Figura 25. Visualización de hallazgos

Fuente: elaboración propia

A continuación, se observa el monto por vendedor, y las líneas de colores representan el medio de contacto más eficiente para cerrar una venta: en primera instancia, WhatsApp; luego, la página web y las llamadas. Por tanto, se entiende que estas son las mejores formas de atraer alumnos y conseguir una mayor cantidad de matriculados. Asimismo, se muestran los vendedores que contactan a más alumnos: Julissa, en primer lugar; seguida de Andrea y Cristina, en segundo y tercer lugar, respectivamente.

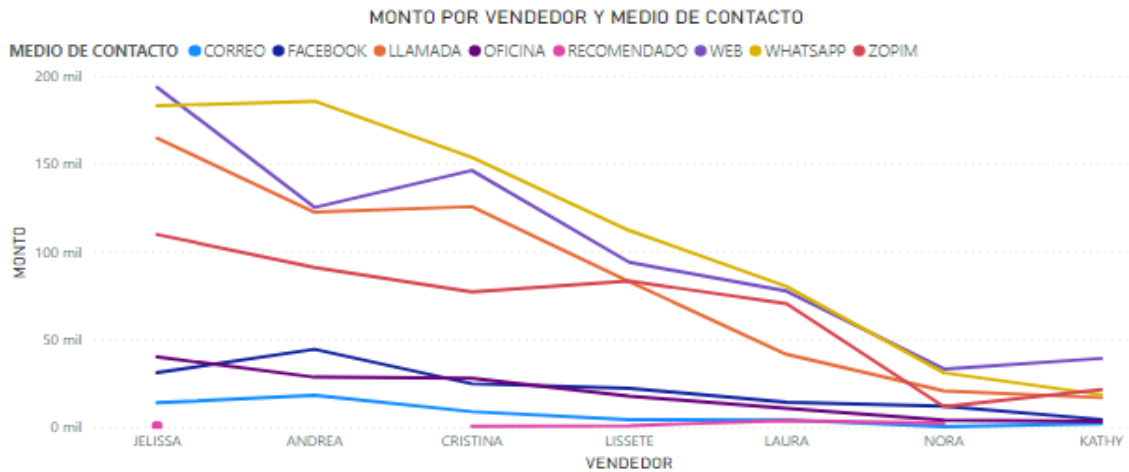


Figura 26. Monto por vendedor y medio de contacto

Fuente: elaboración propia

En el siguiente gráfico, se muestran los seis cursos más demandados del CCDG, donde el más concurrente es el de certificación OSCE, seguido de los diplomas de SIAF, SIGA y Seace, según el monto facturado.

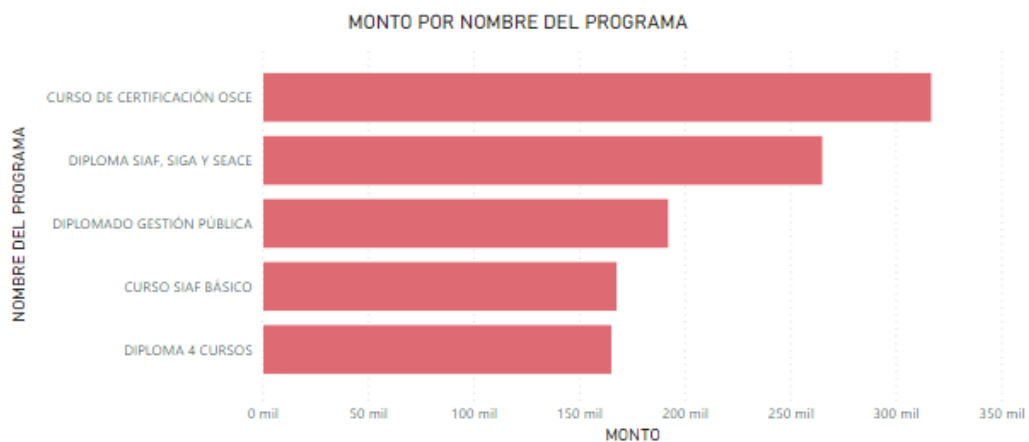


Figura 27. Monto por nombre del programa

Fuente: elaboración propia

En este gráfico, se puede ver que los cursos y diplomas son los más demandados por los estudiantes. Tal dato permite abrir más vacantes para los alumnos en estos programas, porque son los que mayor demanda podrían tener; así, se necesitaría una mayor cantidad de cupos, a fin de cubrir la demanda proyectada.

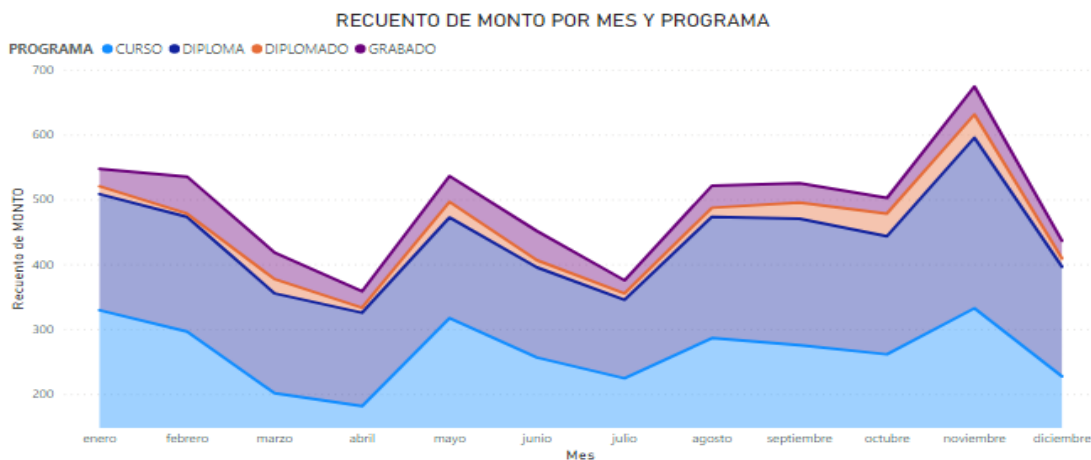


Figura 28. Monto por mes y programa

Fuente: elaboración propia

8.3 Estrategia de implementación

La investigación del presente trabajo podría ayudar al CCDG a proyectar las ventas de acuerdo con las variables que se consideren prudentes para incluirlas dentro de la data. Según los hallazgos, se recomiendan las siguientes estrategias, las cuales deben ser implementadas en la empresa por sus colaboradores:

- Captar a los posibles nuevos clientes, utilizando los medios de contacto más efectivos según el análisis de datos obtenido.
- Capacitar al personal para que pueda recolectar datos de calidad de los clientes, de manera veraz, segura y confiable dentro de la plataforma de la empresa.

- Contar con el apoyo de personal capacitado que se adecúe al método de enseñanza más demandado por los alumnos. Asimismo, se debe contar con evaluaciones constantes de la plana docente, de acuerdo con las encuestas que se realizan a los alumnos.

8.4 Calibrado e implementación

8.4.1 Métodos de ajuste

Para este punto del calibrado, se tomó el 60 % de la muestra, con el objetivo de determinar el modelo de regresión; seguidamente, se tomó el 40 % para generar las pruebas, a fin de consignar la calibración del modelo presentado.

Datos:

Población = 5890

Tamaño de la muestra = 545

545 = 100 %

60 % = 327 - Con esta cantidad de registros se realiza la regresión.

40 % = 218 - Con esta cantidad de registros se efectúan las pruebas de calibración.

n = 327

Fórmula = f (N° de alumnos matriculados); proyectado = 0,7559 (mes) + 4,1642

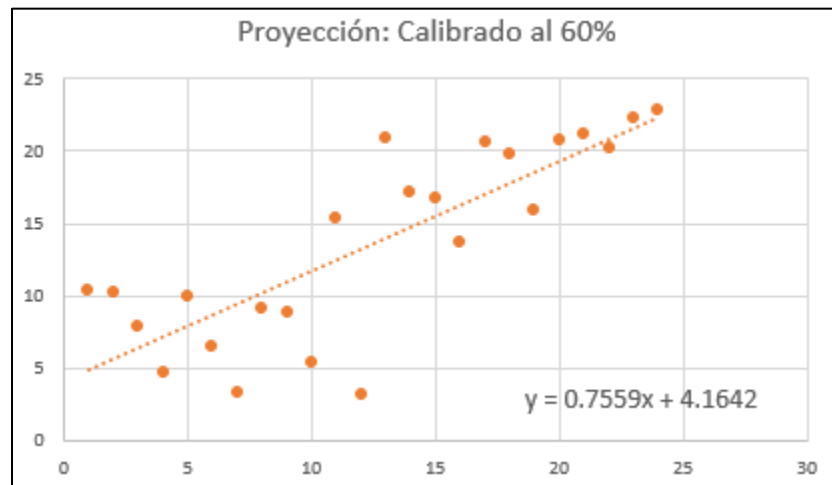


Figura 29. Demanda proyectada

Fuente: elaboración propia

Calibrando el modelo al 40 %

Datos:

Población = 5890

Tamaño de la muestra = 545

545 = 100 %

60 % = 327 - Con esta cantidad de registros se realiza la regresión.

40 % = 218 - Con esta cantidad de registros se efectúan las pruebas de calibración.

n = 218

Fórmula = f (N° de alumnos matriculados); proyectado = 0,5039 (mes) + 2,7761

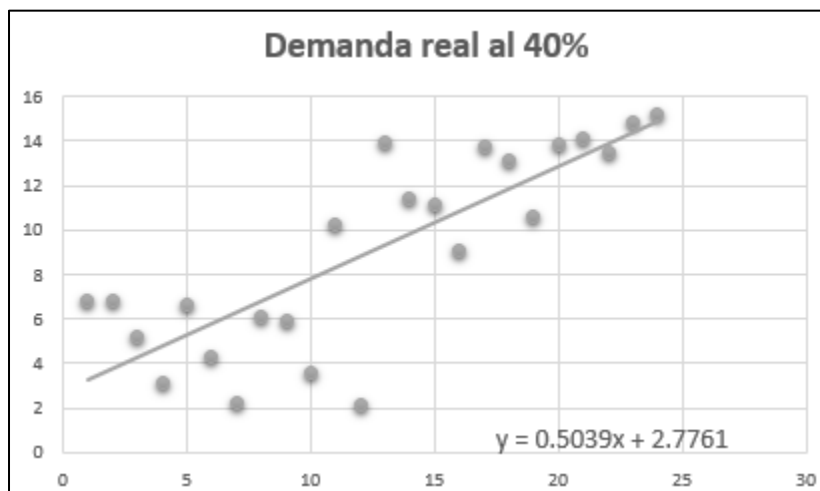


Figura 30. Demanda real al 40 %

Fuente: elaboración propia

Meses(X)	60%	40%	Y= 0.7559X+4.1642	Y=0.5039X+2.7761	EMP	EMC	DEE	DAM	PEMA (%)
1	10	7	4.9201	3.28	1.640	2.69	8.398	1.640	33%
2	10	7	5.676	3.7839	1.892	3.58	7.001	1.892	33%
3	8	5	6.4319	4.2878	2.144	4.60	5.731	2.144	33%
4	5	3	7.1878	4.7917	2.396	5.74	4.588	2.396	33%
5	10	7	7.9437	5.2956	2.648	7.01	3.572	2.648	33%
6	7	4	8.6996	5.7995	2.900	8.41	2.683	2.900	33%
7	3	2	9.4555	6.3034	3.152	9.94	1.921	3.152	33%
8	9	6	10.2114	6.8073	3.404	11.59	1.286	3.404	33%
9	9	6	10.9673	7.3112	3.656	13.37	0.778	3.656	33%
10	5	4	11.7232	7.8151	3.908	15.27	0.397	3.908	33%
11	15	10	12.4791	8.319	4.160	17.31	0.143	4.160	33%
12	3	2	13.235	8.8229	4.412	19.47	0.016	4.412	33%
13	21	14	13.9909	9.3268	4.664	21.75	0.016	4.664	33%
14	17	11	14.7468	9.8307	4.916	24.17	0.143	4.916	33%
15	17	11	15.5027	10.3346	5.168	26.71	0.397	5.168	33%
16	14	9	16.2586	10.8385	5.420	29.38	0.778	5.420	33%
17	21	14	17.0145	11.3424	5.672	32.17	1.286	5.672	33%
18	20	13	17.7704	11.8463	5.924	35.09	1.921	5.924	33%
19	16	11	18.5263	12.3502	6.176	38.14	2.683	6.176	33%
20	21	14	19.2822	12.8541	6.428	41.32	3.572	6.428	33%
21	21	14	20.0381	13.358	6.680	44.62	4.588	6.680	33%
22	20	13	20.794	13.8619	6.932	48.05	5.731	6.932	33%
23	22	15	21.5499	14.3658	7.184	51.61	7.001	7.184	33%
24	23	15	22.3058	14.8697	7.436	55.30	8.398	7.436	33%
327					4.538	23.637	1.782	4.53810	33.00%

Figura 31. Margen de error

Fuente: elaboración propia

8.5 Aplicando la calibración del modelo

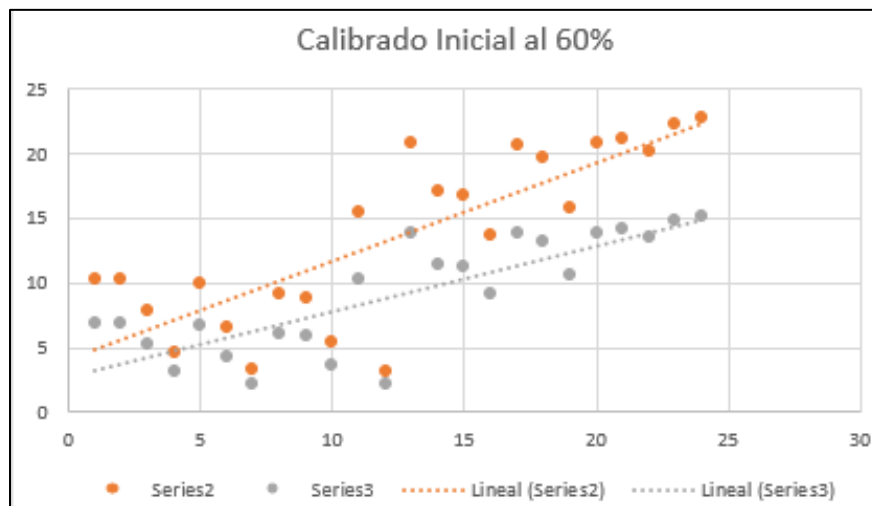


Figura 32. Calibración

Fuente: elaboración propia

Interpretación

Se observa en el gráfico que la línea de tendencia real (línea azul) se encuentra medianamente distante de la línea de tendencia proyectada (línea naranja); por lo tanto, el modelo no es muy exacto en relación con el pronóstico de alumnos matriculados.

9 RETROALIMENTACIÓN

Para este punto, se decidió probar el ajuste del porcentaje del modelo en relación con la muestra, y se determinó que el porcentaje asignado sería del 80 % de la muestra (545). Con esto, se procedió a calcular una regresión con base en el nuevo porcentaje y, sobre ello, los márgenes de errores expuestos en la tabla a continuación.

Total de la muestra = 545

70 % de la muestra $n = 327$

Fórmula = $f(N^\circ \text{ de alumnos inscritos}); \text{proyectado} = 0,8819(N^\circ \text{ de mes}) + 4,8582$

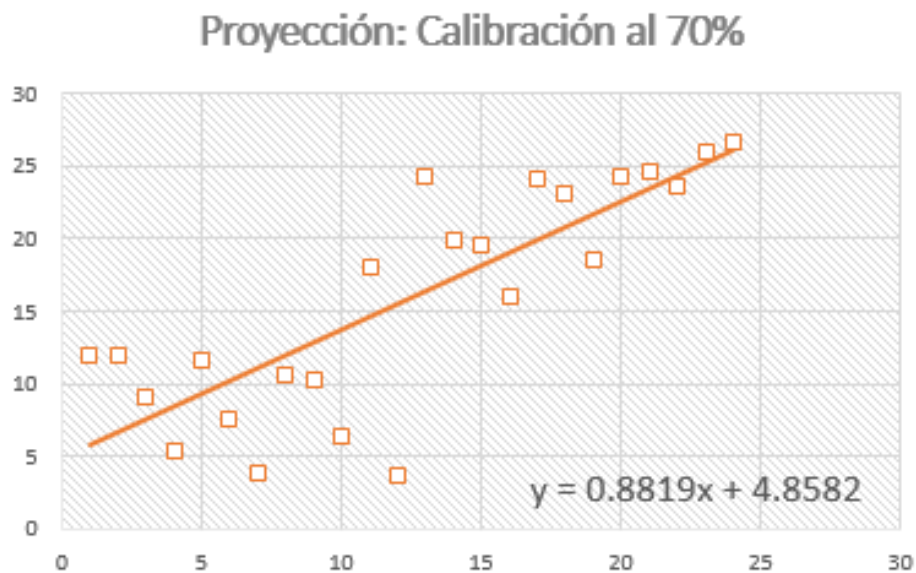


Figura 33. Demanda proyectada al 70 %

Fuente: elaboración propia

Total de la muestra = 545

30 % de la muestra $n = 163$

Fórmula = f (N° de alumnos inscritos); proyectado = $0,3779x + 2,8582$ (N° de meses)

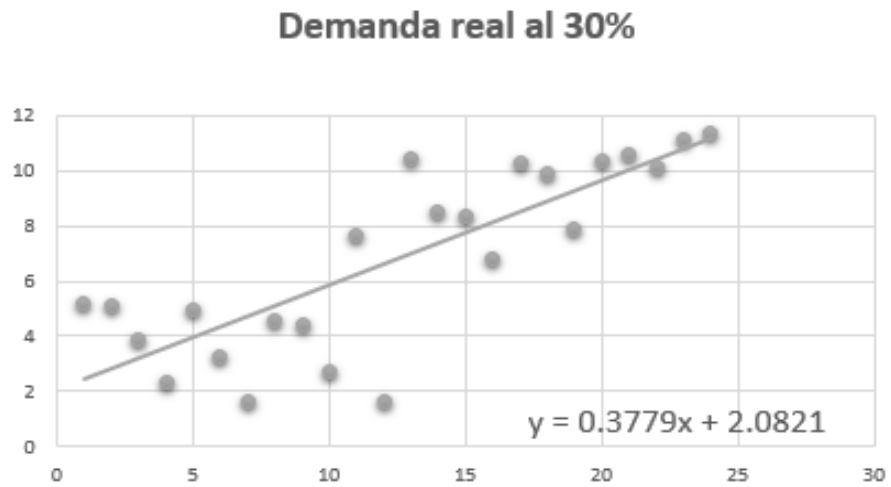


Figura 34. Demanda real al 30 %

Fuente: elaboración propia

9.1 Calibrando la regresión calculada

Meses(X)	70%	30%	$Y = 0.8819X + 4.8582$	$Y = 0.3779X + 2.0821$	EMP	EMC	DEE	DAM	PEMA (%)
1	12	5	5.7401	2.46	3.280	10.76	33.594	3.280	13%
2	12	5	6.622	2.8379	3.784	14.32	28.005	3.784	13%
3	9	4	7.5039	3.2158	4.288	18.39	22.925	4.288	13%
4	5	2	8.3858	3.5937	4.792	22.96	18.353	4.792	13%
5	12	5	9.2677	3.9716	5.296	28.05	14.288	5.296	13%
6	8	3	10.1496	4.3495	5.800	33.64	10.732	5.800	13%
7	4	2	11.0315	4.7274	6.304	39.74	7.684	6.304	13%
8	11	5	11.9134	5.1053	6.808	46.35	5.144	6.808	13%
9	10	4	12.7953	5.4832	7.312	53.47	3.112	7.312	13%
10	6	3	13.6772	5.8611	7.816	61.09	1.588	7.816	13%
11	18	8	14.5591	6.239	8.320	69.22	0.572	8.320	13%
12	4	2	15.441	6.6169	8.824	77.86	0.064	8.824	13%
13	24	10	16.3229	6.9948	9.328	87.01	0.064	9.328	13%
14	20	9	17.2048	7.3727	9.832	96.67	0.572	9.832	13%
15	20	8	18.0867	7.7506	10.336	106.83	1.588	10.336	13%
16	16	7	18.9686	8.1285	10.840	117.51	3.112	10.840	13%
17	24	10	19.8505	8.5064	11.344	128.69	5.144	11.344	13%
18	23	10	20.7324	8.8843	11.848	140.38	7.684	11.848	13%
19	19	8	21.6143	9.2622	12.352	152.57	10.732	12.352	13%
20	24	10	22.4962	9.6401	12.856	165.28	14.288	12.856	13%
21	25	11	23.3781	10.018	13.360	178.49	18.353	13.360	13%
22	24	10	24.26	10.3959	13.864	192.21	22.925	13.864	13%
23	26	11	25.1419	10.7738	14.368	206.44	28.005	14.368	13%
24	27	11	26.0238	11.1517	14.872	221.18	33.594	14.872	13%
381					9.076	94.547	3.564	9.07610	13.00%

Figura 35. Margen de error

Fuente: elaboración propia

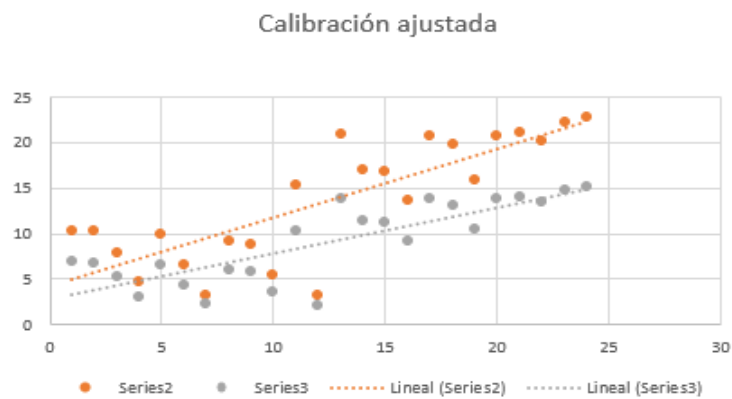


Figura 36. Calibración ajustada

Fuente: elaboración propia

Fórmulas aplicadas:

EMP (error medio del pronóstico):

$$EMP = \frac{\sum_{t=1}^n (Y_t - Y'_t)}{n} \quad e_t = Y_t - Y'_t$$

EMC (error medio al cuadrado)

DDE (dispersión de los errores en relación con el EMP)

$$DDE = \sqrt{\frac{\sum_{t=1}^n (e_t - EMP)^2}{n - 1}}$$

DAM (ABS de la media)

$$DAM = \frac{\sum_{t=1}^n |Y_t - Y'_t|}{n}$$

PEMA (porcentaje de error medio absoluto): a mayor porcentaje, mayor error.

$$PEMA = \frac{\sum_{t=1}^n \left| \frac{Y_t - Y'_t}{Y_t} \right| \cdot 100}{n}$$

10 CONCLUSIONES

- Para cubrir la demanda del año 2020, se aplicó una regresión lineal que permitió predecir la cantidad de alumnos posibles matriculados. Se tomaron como referencia los datos históricos del año 2018 y 2019 de la cantidad de alumnos inscritos en esos años. Con ello, se tuvo que, para el presente año, se tendría una cantidad aproximada de 6380 alumnos.
- Para ajustar el modelo de regresión, se trabajó con la muestra obtenida de toda la población, esta es, 545 registros de los alumnos matriculados, los cuales fueron escogidos al azar. Esto permitió demostrar que el modelo tenía un margen de error del 33 %, pues se obtuvo una proyección con el 60 % de los datos y un real con el 40 %. Con este resultado se pudo implementar un ajuste: se tomó un 70 % de los datos para hacer la proyección y un 30 % como dato real; con esto se consiguió un margen de error inferior al 13 %. Todo esto quiere decir que la demanda no está tan sobreestimada y que se puede acercar al dato real.
- Dentro de las variables para captar potenciales alumnos, se concluyó que WhatsApp, la página web y las llamadas telefónicas están dentro de los mejores medios de contacto. Por tanto, es importante informar a los asesores comerciales que, preferiblemente, deben utilizar estos con los interesados, a fin de obtener una mayor cantidad de matriculados.
- Se identificó que el curso con mayor demanda entre los alumnos matriculados es el de certificación de OSCE, seguido del diploma de SIAF.
- El CCDG cuenta con métodos de enseñanza que se adecúan al alumno; sin embargo, el más demandado es el virtual, debido a la coyuntura de la nueva normalidad a causa de la pandemia.

11 RECOMENDACIONES

- Según la Figura 17, donde se muestra el porcentaje por modalidad de estudio, se determinó que los alumnos prefieren la modalidad virtual. Se recomienda brindar un descuento para obtener una mayor atención por los alumnos, además de contar con una plana docente preparada y con los implementos necesarios para dictar las clases o beneficios adicionales a los matriculados para incentivar su aprendizaje.
- Según la Figura 18, con título Monto por programa, se puede observar que hay una mayor demanda de los cursos y diplomas; por lo cual se recomienda dar apertura a una mayor cantidad de vacantes para esos dos programas.
- Según la Figura 20, con título Monto por medio de contacto, se recomienda que los principales medios de captación de clientes sean a través de WhatsApp y llamadas telefónicas, a fin de incrementar las ventas.
- Según la Figura 21, con título Monto por vendedor, las mejores vendedoras son Jelissa y Andrea; por lo cual, se recomienda darles algún tipo de incentivo para que mantengan el nivel de las ventas.
- Según la Figura 23, en donde se muestra la regresión lineal proyectada para el año 2020, se propone considerar un total de 6380 vacantes para cubrir la demanda.
- Según la Figura 22, se recomienda abrir una mayor cantidad de cupos en los primeros cinco cursos con más preferencia entre el alumnado.

12 FACTORES CRÍTICOS DE ÉXITO PARA EL PROYECTO

- Previamente al diseño de los gráficos de visualización para su análisis, es de suma relevancia contar con una data limpia y depurada, con el propósito de tener información confiable.
- Un factor crítico para el proyecto es la elección de una herramienta adecuada para la visualización de la data y su respectivo análisis, para que ello contribuya a responder las preguntas de investigación.
- Es de vital importancia para el éxito del proyecto contar con la data adecuada y seleccionarla de manera que ayude a cumplir con los objetivos del proyecto.

13 NUEVAS PREGUNTAS

- ¿Existe algún tipo de estacionalidad con respecto a la demanda de los programas de capacitación?
- Según la proyección realizada en la Figura 19, se obtuvo un total de 6380 posibles alumnos matriculados. Del total, ¿cómo es la distribución de la cantidad de alumnos por programas?
- ¿De qué manera se puede reducir el margen de error de la proyección para que sea menor al 13 %?

14 DECISIONES A TOMAR A PARTIR DE LOS DATOS ESTUDIADOS

- Dar lugar a más vacantes en los cursos de OSCE y en los diplomas de SIAF, SIGA y Seace, dado que son los más demandados, según la Figura 22.
- Según la proyección mostrada en la Figura 19, se pueden conseguir 6380 vacantes para el año 2020.
- Se debe hacer un reconocimiento e incentivo a las vendedoras más destacadas durante cada trimestre, de forma que se vean estimuladas las ventas.
- De acuerdo con la Figura 13, donde se identificaron los medios de contacto más efectivos, deben brindarse celulares con planes empresariales a los vendedores, de forma que puedan mantener contacto con los alumnos potenciales a través de WhatsApp y llamadas.

15 BIBLIOGRAFÍA

- CEPEG. (s.f.). *Centro Peruano de Estudios Gubernamentales*.
<https://www.cepeg.pe/quienes-somos>
- Eucariota. (s.f.). *Introducción a Power BI*. <https://eucariota.eu/curso/curso-power-bi/>
- IBM Analytics. (s.f.). *Metodología fundamental para la ciencia de datos*.
<https://www.ibm.com/downloads/cas/6RZMKDN8>
- IPAPPG. (s.f.). *¿Quiénes somos?* Obtenido de <https://ipappg.edu.pe/nosotros.html>
- LumenSoft. (2018). *Seace 2018: estadísticas de licitaciones del Estado*.
<https://www.lumensoft.pe/seace/seace-2018/>
- Ochoa, C. (2015). *Muestreo probabilístico o no probabilístico*.
<https://www.netquest.com/blog/es/blog/es/muestreo-probabilistico-o-no-probabilistico-ii>
- ONU Mujeres. (2012). *Análisis PESTEL*. <https://www.endvawnow.org/es/articles/1182-analisis-pestel.html>
- Oyelowo, A. (2018). *Introduction to Power BI*. <https://foresightbi.com.ng/microsoft-power-bi/introduction-to-powerbi/>
- PowerData. (s.f.). *Calidad de datos. Cómo impulsar tu negocio con los datos*.
<https://www.powerdata.es/calidad-de-datos>
- QuestionPro. (s.f.). *Calculadora de muestras*. <https://www.questionpro.com/es/calculadora-de-muestra.html>
- R&C Consulting. (s.f.). *Home*. <https://rc-consulting.org/cursos-presenciales/>
- UGR. (s.f.). *Tema 6: distribuciones muestrales*. <https://www.ugr.es/~mvargas/tema6sd.pdf>

Universidad de Alcalá. (s.f.). *Data Science certificate programs online*. <https://www.master-data-scientist.com/data-science-certificate-programs-online>

16 ANEXOS

Anexo 1. Cursos, Diplomas y Diplomados

Cursos: los cursos que se dictan son cortos, tienen una duración entre 4 a 8 días. En cuanto a los temas son diversos como Administración, gestión pública, contabilidad, logística, entre otros. Tienen un costo promedio entre 400 y 500 soles.

Sistema Integrado de Gestión Administrativa - SIGA- Logístico y Patrimonio

Sistema electrónico de contrataciones del estado - SEACE 3.0

Planeamiento estratégico y gestión por resultados

Sistema integrado de gestión administrativa - SIAF RP

Gestión de contrataciones del estado|

Certificación por niveles del OSCE

Gestión pública e Inverte.pe

Contabilidad gubernamental

Gestión de tesorería

Diploma: tienen una duración entre 20 A 50 días de capacitación y un costo entre los 500.00 y 1,500.00 soles.

SIAF SP

SIGA MEF

SIAF SP Y SIGA MEF

SIAF SP, SIGA MEF Y SEACE

Finanzas Públicas

Gestión De Contrataciones Del Estado

Diplomado: tienen una duración mínima de 7 meses, 384 horas lectivas y 24 créditos académicos

Gestión Pública

Modernización del Estado

Administración y Gestión Pública

Sistemas Administrativos del Estado