San Jose State University
SJSU ScholarWorks

Master's Projects

Master's Theses and Graduate Research

Spring 2012

Visualization of Online Deals

Praveen Thiruvathilingam
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_projects

Part of the Computer Sciences Commons

Recommended Citation

Visualization of Online Deals

A Project Report

Presented to

The Faculty of the Department of Computer Science

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Computer Science

by

Praveen Thiruvathilingam

May 2012

SAN JOSE STATE UNIVERSITY

The Undersigned Project Committee Approves the Project Titled

Visualization of Online Deals

by
Praveen Thiruvathilingam

APPROVED FOR THE DEPARTMENT OF COMPUTER SCIENCE
SAN JOSÉ STATE UNIVERSITY

May 2012

Dr. Soon Tee Teoh, Department of Computer Science   Date

Dr. Robert Chun, Department of Computer Science   Date

Prof. Frank Butt, Department of Computer Science   Date

APPROVED FOR THE UNIVERSITY

Associate Dean Office of Graduate Studies and Research   Date

ABSTRACT

Visualization of Online Deals

by Praveen Thiruvathilingam

This project identifies the impact of online deals and coupons on the life of people. The basic idea is to find the trend in the sales of these deals and this project would be helpful for companies, restaurants and dealers who are trying to sell coupons to popularize their product. The final output of the project would be a timeline graph with the deals displayed based on the month of their sale. Information like the original price, deal price, discount and number of coupons sold will be displayed in a pop-up window when a deal is selected. Different colors are used to differentiate the different types of deals. A pie chart depicting this information for various US states is created which provides a summarized view of the deals. This way, an overall picture about the deals sold for a particular month can be obtained.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Soon Tee Teoh for his guidance, valuable insight and extensive support. Moreover, I would like to thank my committee members Dr. Robert Chun and Prof. Frank Butt for their critique and support.

Finally, I want to thank my parents and friends for their encouragement and moral support.

<center>TABLE OF CONTENTS</center>

List of Tables

List of Figures

1. Introduction

People look for best deals all the time when they want to purchase something. "One deal one day" or "flash deals" as they are called is a type of ecommerce business established initially by Amazon and eBay. The members who have signed up for alerts get messages or emails daily on deals that are available for the day. This business sprung and gained popularity in 2009 and became a big sensation by 2010. A number of startups whose business is primarily selling these deals came up and Groupon is one among them. Currently, Groupon has revenue of over $300 million [8].

We have seen that a variety of deals are available, like electronics deals, movie coupons, travel deals, restaurant deals, automobile deals and various types of entertainment coupons. Not all the deals are popular during all the months. There is a trend which exists and if followed, can increase the sales of these deals. In other words, we need to understand the purchasing psychology of people. For example, we see that there are many electronics deals during Thanksgiving weekend and many travel related deals during long weekend. So how do sellers decide which month they need to sell their products and which cities they need to concentrate? This is a big research topic in many universities across the globe.

We can find out the trend in deals sales by analyzing the history. So the idea behind this project is to come up with a graphical visualization for online deals. We have chosen to use the historical deals available from the popular website Groupon.com. We have collected data for 25 US cities and have developed an interactive timeline graph which

could be used to analyze and understand the trend behind these deals for these cities i.e. to identify the type of deals that are preferred during various months. To make the graph more informative, we have provided different colors for different types of coupons. The width of the rectangular bar representing the deals is relative to the number of coupons sold.

A pie chart depicting the overall information for each state is created and the pie charts are placed on a US map so that a comprehensive view for the entire country can be obtained at a glance.

1.1 How Deals work

The deals profit both the retailer and the consumers. This is one of the biggest advantages of the deals that make it popular. The consumers enjoy products while the retailers are able to sell their surplus inventory. Few retailers give away the products to popularize the product. Majority of these deals are local businesses thus concentrating a particular city or region where the product can be sold at large. Retailers buy the products in groups and sell the product at a price below their RRP. Statistics have shown that retailers are more profitable than imagined [8]. This can be very well understood from the Groupon revenue which has increased by many folds over the last 2 years.

When a deal goes live, the subscribers are alerted about the deals through emails and messages. These deals are not mere sales. They are a type of marketing strategy. Many
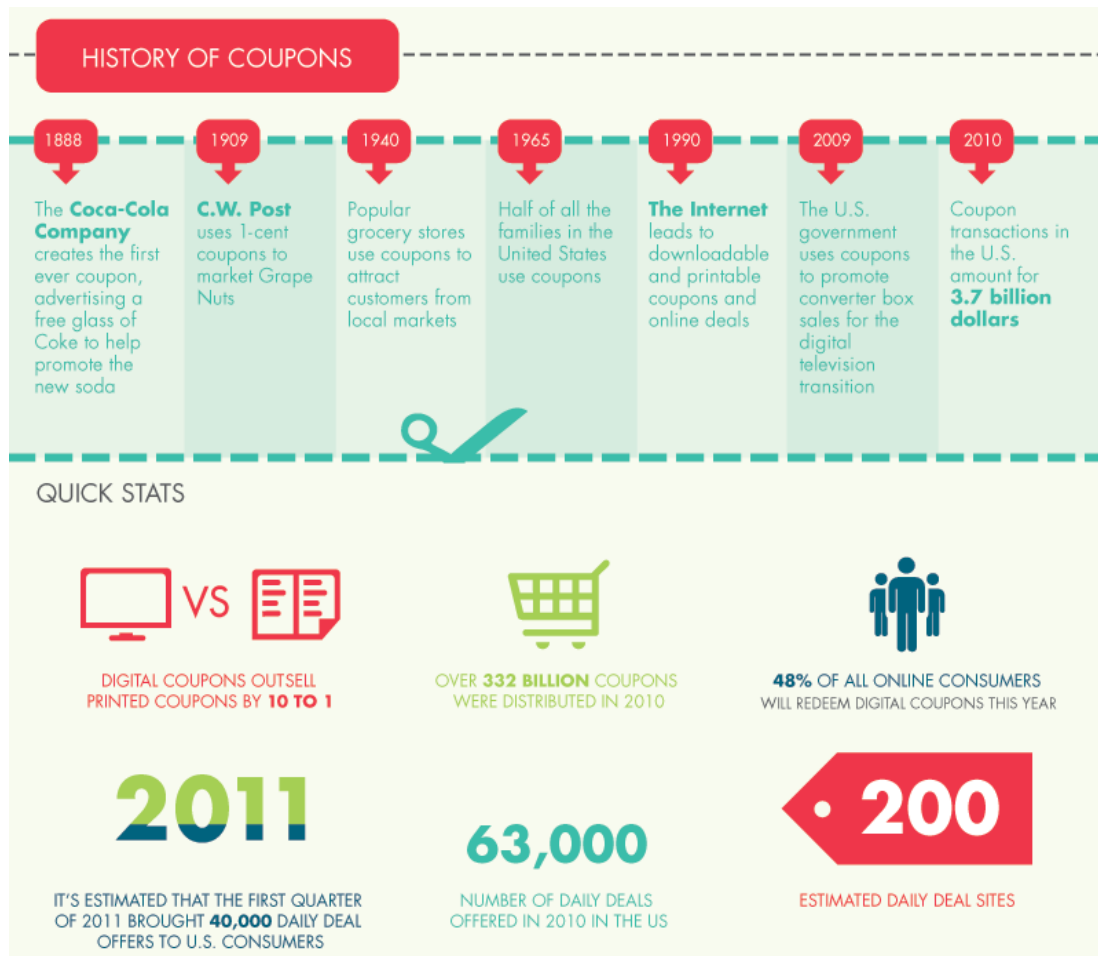
Figure 1- History of coupons

companies, restaurants, hotels, auto deals and small business gain exposure from this. Thus deals help small businesses to stabilize in a particular region. Business reports have shown that smaller businesses have shown 140% increase in sales over six months period. So the primary purpose of deals is ecommerce but it serves as advertisement and marketing platform as well. A small study reveals that marketing via deals accounts for

23.5% of a company's annual spending while email promotion and Google ad works are

16.1% and 14.7% respectively [18].

HISTORY OF ONLINE DEALS

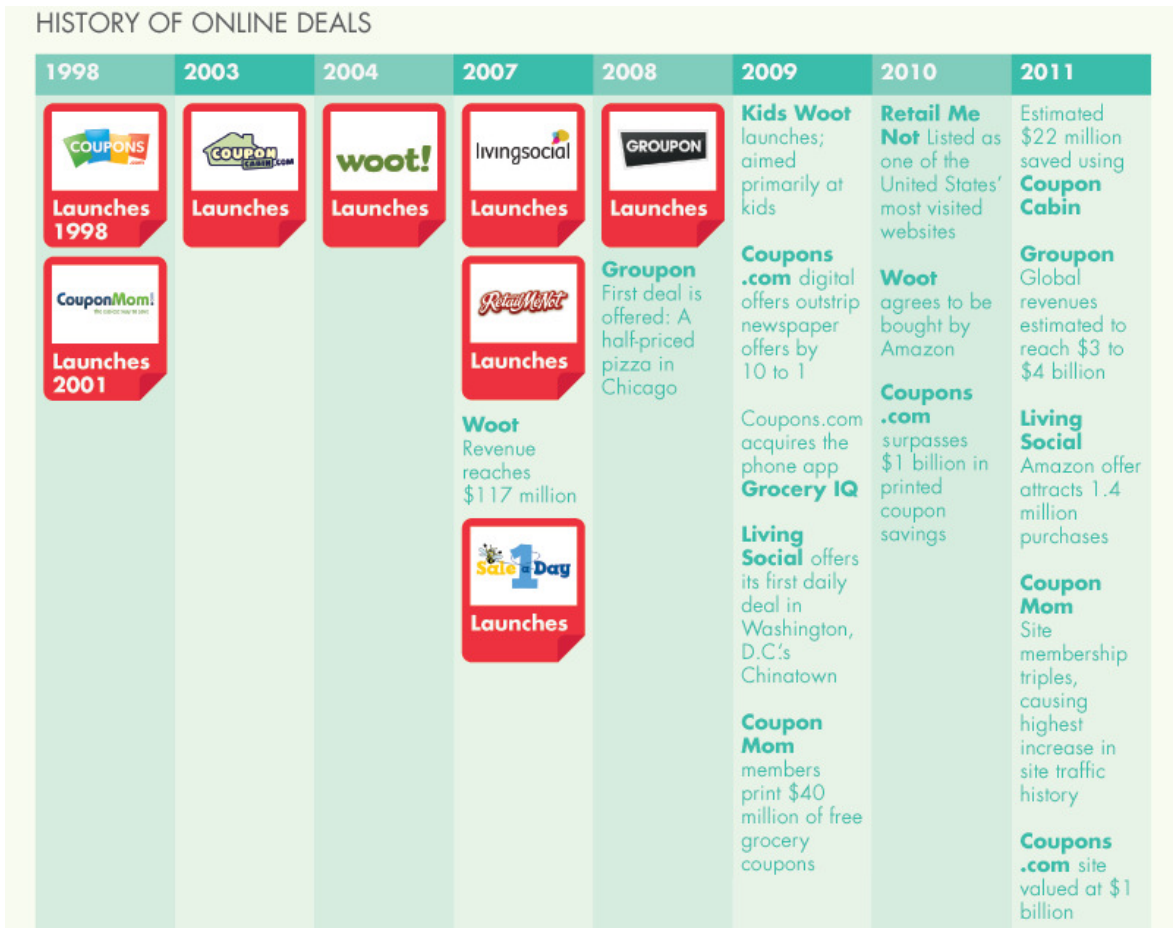| 1998	2003	2004	2007	2008	2009	2010	2011
COUPONS **Launches 1998**	COUPON CABIN.com **Launches**	woot! **Launches**	livingsocial **Launches**	GROUPON **Launches**	**Kids Woot** launches; aimed primarily at kids	**Retail Me Not** Listed as one of the United States' most visited websites	Estimated $22 million saved using **Coupon Cabin**
CouponMom! **Launches 2001**			RetailMeNot **Launches**	**Groupon** First deal is offered: A half-priced pizza in Chicago	**Coupons .com** digital offers outstrip newspaper offers by 10 to 1	**Woot** agrees to be bought by Amazon	**Groupon** Global revenues estimated to reach $3 to $4 billion
			Woot Revenue reaches $117 million		Coupons.com acquires the phone app **Grocery IQ**	**Coupons .com** surpasses $1 billion in printed coupon savings	**Living Social** Amazon offer attracts 1.4 million purchases
			Sale a Day **Launches**		**Living Social** offers its first daily deal in Washington, D.C.'s Chinatown		**Coupon Mom** Site membership triples, causing highest increase in site traffic history
					Coupon Mom members print $40 million of free grocery coupons		**Coupons .com** site valued at $1 billion

Figure 2 – Evolution of Online deals

Figure 1 shows the history of coupons and also various statistics related to it. We can see

that digital deals are quickly replacing the printed coupons that had existed over the past

100 years. From Figure 2, it is evident that online deals are a potentially huge business
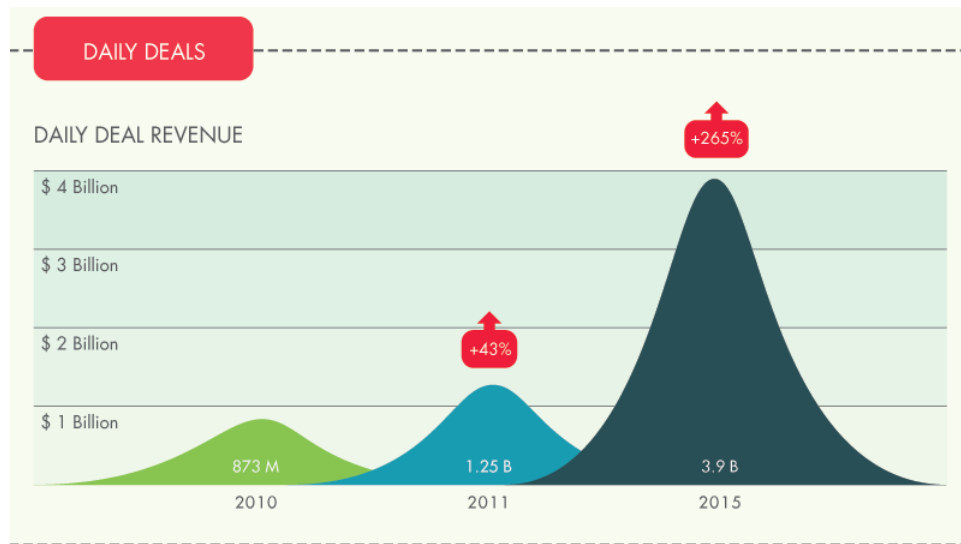
with revenue crossing $22 Billion.

Figure 3 – Revenue chart for the online deals

Figure 3 shows that there would be 265% increase in online deals sales by 2015 which is more than 3 times the overall revenue in 2011. Some of the common products sold are tickets for special events; Tourism related services, Health and fitness, Auto services, Restaurant and Bars and medical services.

2. Related Works

2.1 Introduction to Related Works

Because of the popularity of the online deals, a number of researchers around the globe are tracking the websites selling the deals to identify the trend and to understand the purchasing psychology of the people. Numerous domestic and international journals have been read before forming the basis of this project.

2.2 Fundamentals Concepts

The paper titled, "Mining Online Deal Forum for Hot Deals" provided ideas on mining information regarding deals that are shared by people in forums and blogs. People tend to buy deals based on deal alerts via email from the websites. In this article, the authors discussed the various ways of interpreting the messages in forums to identify the deals that are "hot" and were bought in large numbers. Mechanism to intelligently identify seasonal deals was also discussed. This paper proved useful to my project because I am trying to identify and represent all the deals in a graphical and interactive manner [1].

2.3 e-coupon Studies

The paper entitled, "Effects of perceived behavioral control on the consumer usage intention of e-coupons" recorded the changes e-coupons could bring on to the lives of consumers. A model was created to estimate the coupon usage behavior of the consumers. An empirical study with 588 consumers was made. The model also threw light on the internet usage behavior of the people who look for the best coupon to suit their needs. This study proved useful to my project as I will be trying to demonstrate a similar behavior of people in a graphical way where I will highlight the number of coupons sold every month [11].

The paper entitled, "Price discrimination through online couponing: Impact on likelihood of purchase and profitability" studies the satisfying effects of online deals. One major result I am going to reveal through my project is the profit that both the seller and buyer get using these deals. This paper explains the importance of coupons in bringing down

the price so that profit disparity that exist between the seller and buyer is reduced by a large extend which in turn helps in stabilizing the market [12].

2.4 Data Mining

Websites are mined on daily basis by companies to know the popularity of their product among people. Largely this is done for advertisement, finding hits, tracking IP address and identifying trends. This information provides insight into the behavior of people. As far as mining from websites is concerned, data mining tools look for keywords and try to find relationships between the keywords and location or zip code. Data mining further goes into classification, clustering and finding association between the extracted information. In most of the times, data mining algorithms look for patterns to match in order to validate the data.

2.5 Data Visualization

Interactive timeline review provides mechanism for treating timelines as structured documents which allows display of metadata information to the events. By placing the events along the one dimensional timeline we are providing a more interactive graphical presentation of data which could be well understood than non graphical representations like tables. A timeline has theme and the events are related to the theme thus this is similar to a video presentation [14].

2.6 2D Matrix Representation

Visualization of online deals has always been a research topic ever since it gained popularity in 2009. Many researchers have come up with various visualizations to present their idea of rendering the deals details in an interactive way. Figure 4 is one such attempt made by a student that got published in "The Harvard Business Review" magazine. They have used 2D bin algorithm to represent the information in a 2D matrix format [13]. According to us, the visualization could have made better if a number of changes were implemented in it. Our project is inspired by this visualization but with plenty of changes and interactive features that makes the visualization more informative.
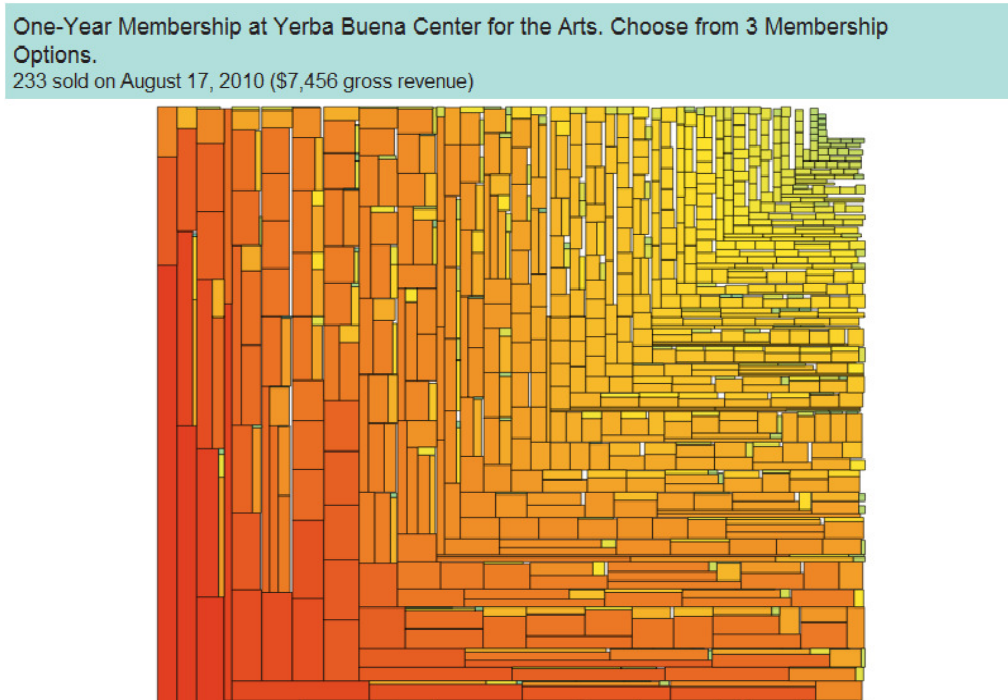

Figure 4 - 2D bin method of visualizing the deals details

2.7 Timeline Chart

Timeline is a way to display the events as they happen in a chronological order. Timeline charts normally consist of horizontal bars with dates and events will be plotted in the timeline area above the bar. For the sake of providing interactivity, zoom in/out bar is provided. This helps the user to view more detailed information by zooming in or just get a glance by zooming out.

Timeline are normally used to represent historical events which would help people understand history better by this graph rather than reading them in a textual format. Timeline can take any scale of time depending upon the need. Historical events can be represented based on years whereas September 11, 2001 events need to be represented on a minute by minute basis.
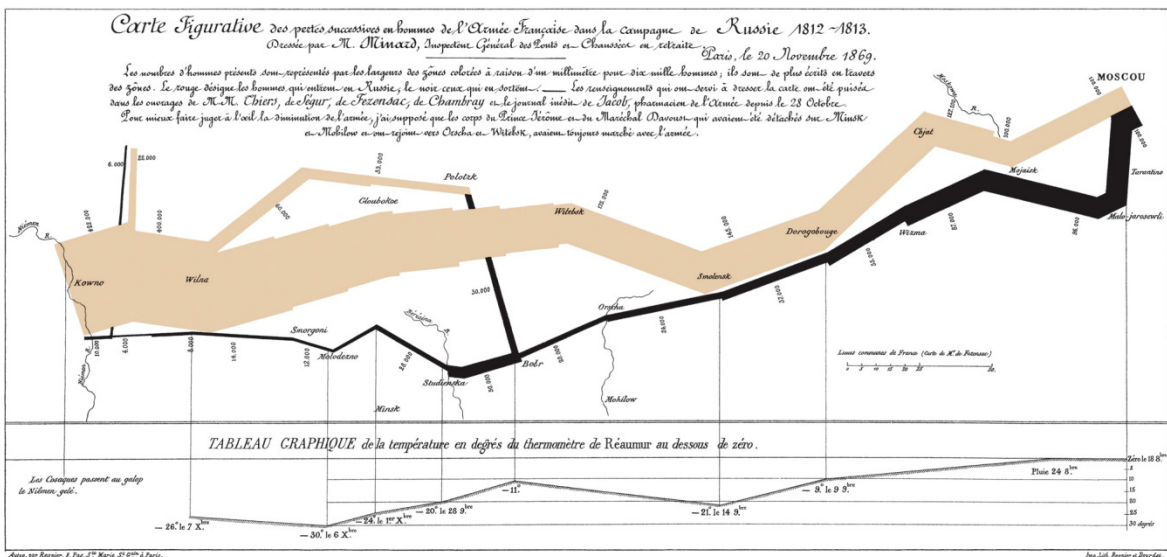

Figure 5 – Napoleon's invasion of Russia

Minard's representation of Napoleon's invasion of Russia is one of the examples that made me choose timeline method for representing deals. In this timeline graph, the author has depicted the path taken by Napoleon to reach Russia. Apart from that, the features to be noticed are

(i) Usage of different colors to represent the path taken from France to Russia and then back to France from Russia

(ii) The width of the bar represents his army strength during the course of his journey

(iii) Various rivers and mountains crossed are represented

The above graph is the best example that I chose for my project. I planned to use the same concept for deals i.e. use of colors and difference in width based on the number of deals sold. The difference in color would help users to identify the deals easily and width would give an idea on the sales of the deal. So if we focus on a particular month, the more prominent color on the screen shows the type of deals that is sold the most that month and the large rectangular bars would show the deals that were sold the most.

3. Design and Implementation

3.1 Data Extraction

For creating the visualization, we need data i.e. history of deals details. This information is not available neither for granted from the websites selling the deals nor are these companies willing to share the information for research purposes. Actually the companies are selling the information for a higher price. So we need to identify other methods of

getting the information. After searching, we found that Groupon has separate pages in Facebook for each city they sell these deals. These pages are updated daily as new deals are posted.

3.1.1 Facebook Graph API

The Facebook Platform provides a set of APIs and tools which enable third party developers to integrate with the "open graph" — whether through applications on Facebook.com or external websites and devices. The Graph API is the core of Facebook Platform, enabling developers to read from and write data into Facebook. The Graph API presents a simple, consistent view of the Facebook social graph, uniformly representing objects in the graph (e.g., people, photos, events, and pages) and the connections between them (e.g., friend relationships, shared content, and photo tags). Facebook authentication is needed to interact with the Graph API. The authentication provides a single sign-on token that can be used across web and mobile apps.

After this authentication is obtained, we can extract information from any user profile, pages, events and activity feed. The information need to be public for us to access these pages else access will be denied. Using this Facebook page access, we extracted the urls for the deals from the relevant city pages. We restricted the information to 3 months. Thus urls leading to Groupon.com deals pages were extracted for 25 US cities. These urls were shortened urls.

Below is a part of the code used for establishing connection with Facebook Graph API and then extracting the url

graphQuery = 'https://graph.facebook.com/grouponbost/feed';
authToken='145634995501895\477bb3c939123a5845afe90d.1100002565213903\F1VA2q2iU6SZX_XXrs'
url = graphQuery+'?date_format=U&limit=1000&access_token='+ authToken +'&callback=?';

3.1.2 Shorten URL

Shortened url is a mechanism to blacklist spam emails. Many websites don't redirect to their website when short urls are used to access them. This happened with the Groupon short url that was extracted. We were not able to reach the desired pages. As a result, a piece of code to recreate the original url from the short url was written. Short urls lead to the same page as the regular urls but their length is substantially short and contains fewer characters but still redirect to the actual website.

Most of the time we won't know the redirection that is taking place. Every short url has the following information which it passes on to the browser "HTTP 30X: Object has moved" HTTP HEADER (which contains the location of the original url)". Now we need to extract the original url from this header. Now the original url are stored in text file and this will serve as the source for our data extraction.

A function written to obtain the original url from shorten url is

function expand_shortlink($url) {
$headers = get_headers($url,1);
if (!empty($headers['Location'])) {
$headers['Location'] = (array) $headers['Location'];
$url = array_pop($headers['Location']);
}

return $url;

}

3.1.3 DOM parsing

After extracting the original url, the next step is to extract the required information by parsing through the page. The information that we are interested are

- Title of the deal

- Date they are posted

- Original price

- Deal price

- Discount

- Number of people who bought the deal

These are the information that is going to appear in the description part when a deal is clicked. These data are going into the JSON feed description part. As we can see, there are many other information that are available in the web page as we parse through but these are the ones that are crucial to our project. The most important part is the number of deals sold as it is directly related to the width of the deals in the timeline chart.

			title	originalprice	amount	discount	amountsaved	couponsold	date
☐	✎	✗	Two or Four Hours of Housecleaning	$180	from$85	53%	$95	Over 53 bought	09/25/2011
☐	✎	✗	$14 for Corn-Maze Outing for Two (Up to $28 value)	$28	$14	50%	$14	Over 470 bought	09/25/2011
☐	✎	✗	If 220 People Donate $10, Then imagine1day Can Bui...	$20	$10	50%	$10	Over 600 bought	09/22/2011
☐	✎	✗	$10 for 8"x8" Photo Book ($29.99 Value) for New Sh...	$30	$10	67%	$20	Over 120 bought	09/20/2011
☐	✎	✗	One, Two, or Four Diamond Microdermabrasion Treatm...	$100	from$49	51%	$51	Over 100 bought	09/26/2011
☐	✎	✗	$35 for Jiffy Lube Signature Service Oil Change, T...	$71	$35	51%	$36	Over 810 bought	09/24/2011
☐	✎	✗	$20 for Two Admissions to Queen City is Haunted To...	$40	$20	50%	$20	Over 610 bought	09/23/2011
☐	✎	✗	Gourmet Pizza and Comfort Fare or Pizza Meal for T...	$20	from$10	50%	$10	Over 300 bought	09/22/2011
☐	✎	✗	$25 for Bicycle Tune-Up ($55 Value)	$55	$25	55%	$30	Over 140 bought	09/22/2011
☐	✎	✗	Media Transfers To DVD. Three Options Available.	$60	from$20	67%	$40	Over 210 bought	09/22/2011
☐	✎	✗	$2 for Five One-Night DVD Rentals ($5 Value)	$5	$2	60%	$3	Over 690 bought	09/21/2011
☐	✎	✗	$49 for Six-Week Baseline Tennis Program and Gold ...	$115	$49	57%	$66	Over 200 bought	09/22/2011
☐	✎	✗	Intro to Pole or 3-Button Strip Tease Workshop	$40	from$20	50%	$20	Over 290 bought	09/21/2011
☐	✎	✗	Brewery History Tour for Two. 14 Options Available...	$60	from$30	50%	$30	Over 60 bought	09/16/2011
☐	✎	✗	$15 for $30 Worth of Jewelry, Gifts, and Asian-ins...	$30	$15	50%	$15	Over 170 bought	09/19/2011
☐	✎	✗	$15 for One-Year Indoor-Go-Karting Membership and ...	$32	$15	53%	$17	Over 500 bought	09/21/2011
☐	✎	✗	One or Three Microcurrent Lunchtime Facelift Treat...	$199	from$49	75%	$150	Over 200 bought	09/21/2011
☐	✎	✗	90-Minute Segway Tour for One or Two People	$60	from$35	42%	$25	Over 1,000 bought	09/19/2011
☐	✎	✗	$5 for One lb. of Double-Chocolate-Chip Fudge Brow...	$10	$5	50%	$5	Over 810 bought	09/18/2011
☐	✎	✗	$99 for Weight-Loss and Detoxification Package Inc...	$350	$99	72%	$251	Over 100 bought	09/19/2011
☐	✎	✗	Appetizers and Entrees for Two or Four	$44	from$22	50%	$22	Over 230 bought	09/15/2011
☐	✎	✗	Therapeutic or Four-Hand Massage Package	$125	from$49	61%	$76	Over 130 bought	09/18/2011

Figure 6 – Table with deals details

By parsing through the DOM tree from the root element, we were able to obtain the above values from their relevant tags. The extracted information is stored in a mysql database tables created individually for the respective cities.
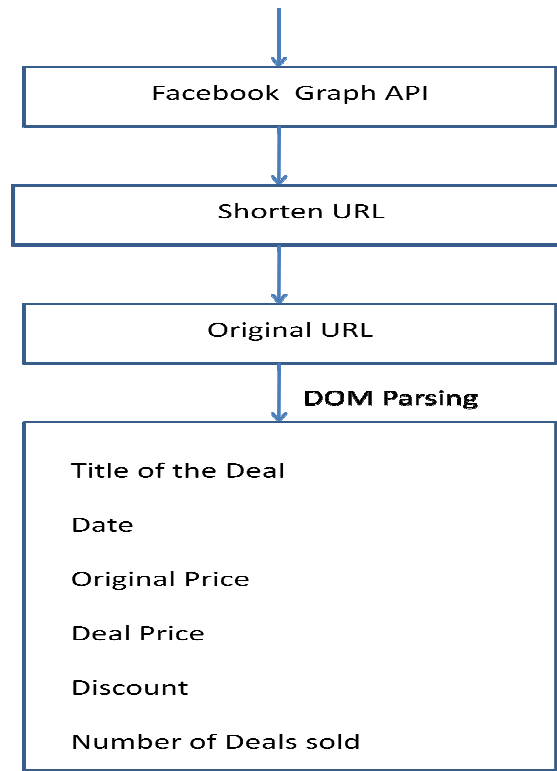
```
                          ┌──────────────────────────────┐
                          │      Facebook  Graph API     │
                          └──────────────────────────────┘
                                         │
                          ┌──────────────────────────────┐
                          │         Shorten URL          │
                          └──────────────────────────────┘
                                         │
                          ┌──────────────────────────────┐
                          │         Original URL         │
                          └──────────────────────────────┘
                                    DOM Parsing
                          ┌──────────────────────────────┐
                          │   Title of the Deal          │
                          │                              │
                          │   Date                       │
                          │                              │
                          │   Original Price             │
                          │                              │
                          │   Deal Price                 │
                          │                              │
                          │   Discount                   │
                          │                              │
                          │   Number of Deals sold       │
                          └──────────────────────────────┘
```

Figure 7 – Steps followed in the data extraction phase

3.2 Timeline Chart

The timeline chart was created by bringing together a number of pieces that makes the graph together. Timeglider is a JQuery plug-in that provided the framework or the basic outline for the chart. The zoom in/out bar was also incorporated with the chart to allow for more interaction. The timeline can be zoomed in by dragging the vertical bar down where more time is shown and zoomed out by dragging the vertical bar up. The zoom in/out can also be achieved by scrolling the mouse wheel. The white area is where the events are represented. The timeline area can be moved left or right to view events accordingly.

3.2.1 JSON

The data for the chart comes in the form of JSON feed. So the data from MySQL has to be converted to JSON format and fed into the chart as input. The outer wrapper represented by square brackets can contain multiple timeline events represented within the curly brackets. Each timeline event is called an OBJECT. Each object has a unique ID, title which is the title of the deal, start date, original price, deal price, discount, Importance and the number of deals sold. The start date would represent deal's start date in the timeline window.

```
[ {
"id": "Deals",

"title": " Atlanta Deals",

"focus_date": "2011-07-01 12:00",

"initial_zoom": "20",

"color": "#82530d",

"events":[
{
"id":"Washington006964",
"title":"$10 for One-Year Magazine Subscription (Up to $28 Value)",
"description":
"Original Price: $28
 Deal Price: $10
Discount: 64%
Amount Saved: $18
Coupons:Over 680 bought",
"startdate":"2011-09-22",
"importance":68
}
```

As we can see in the above code, description contains the data that should be displayed in the pop up window. Importance attribute is proportional to the number of the deals sold.

Importance is calculated as a modulus of the number of deals sold. In the above example, 680 deals were sold hence the importance is 68. For deals sold over 1000, the importance is kept a maximum of 90. The reason for using the JSON format is the necessity to provide certain features needed in the data like the width of the deals rectangular bar being proportional to the number of deals sold and color representing the type of deal.

3.2.2 Timeline Chart Design

In this chart, as we could see the zoom level is 26, and different deals can be seen in various colors. The choice of colors is based on the type of deals. The type of deals classification is based on the standard classification provided in the internet. The width of deals (rectangular bar) is proportional to the number of deals sold. The rectangular bars have a start date similar to the start date given in the JSON input. Since the deals exist for a long time till they are completely sold out, the end date for most deals is not available. The timeframe in the timeline becomes more and more specific as we zoom in along with the deals (events) which get magnified.

As we zoom in/out the focus date gets changed in the view. This will help users to know which timeframe (date) they are viewing. There is only horizontal scrolling and no vertical scrolling.
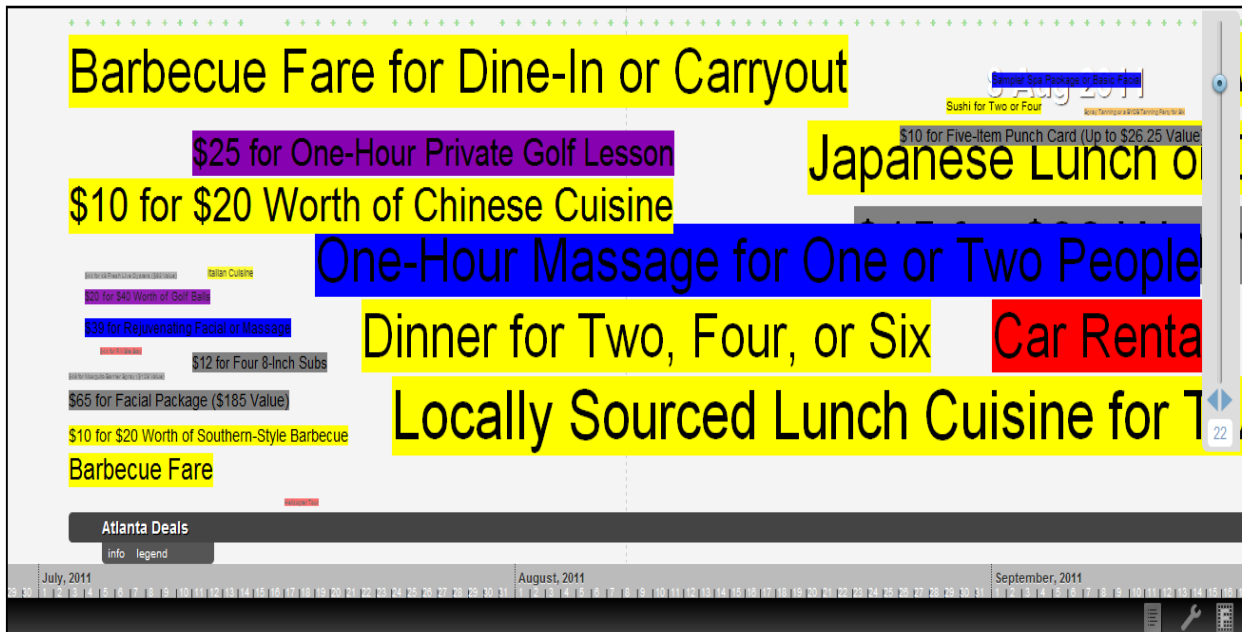
Figure 8 – Trend in deals for the month of July for Atlanta – less deals were sold during

the start of the month and as time passes more deals were sold – primarily food related

deals (Yellow)

Figure 8 is the ideal view. Though we don't get to see the deals with their starting date, we can get an overall picture. As we zoom in, the deals narrow down based on their starting date.

The data available in the description attribute of the JSON feed will be shown when the deals are clicked. A number of deals are posted on the same day. So to avoid moving horizontally, we have made the deals overlap each other. This does not cause any effect on the chart as the deals are of different sizes and hence the possibility of one deal being overlapped / hidden by another is ruled out. Different deals being differentiated by colors add to this. This provides visual difference which is a positive feature.
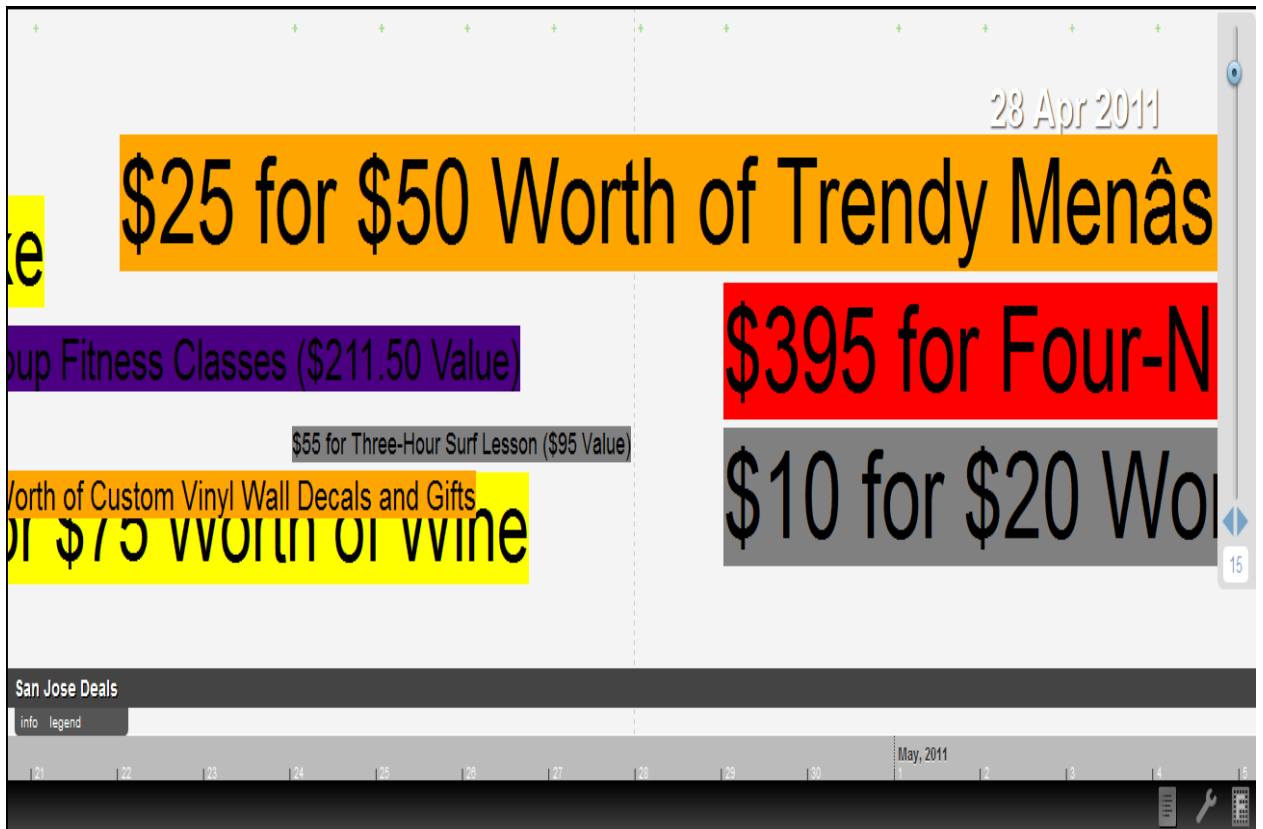
Figure 9 – A more zoomed in view

In figure 9, we could see that the focus date is April 29. The deals that have their starting date closer to April 29 can be seen in a zoomed in view. From the figure, it is evident that 2 deals have their starting date as April 29. The width of the deals is almost the same meaning that equal number of deals were sold.

The next important aspect of this timeline chart is to provide important information about the deals. This is achieved using the pop up window that appears when the deals are clicked. The pop up window opens up to the same width as the timeline chart and shows the information along with the start date.
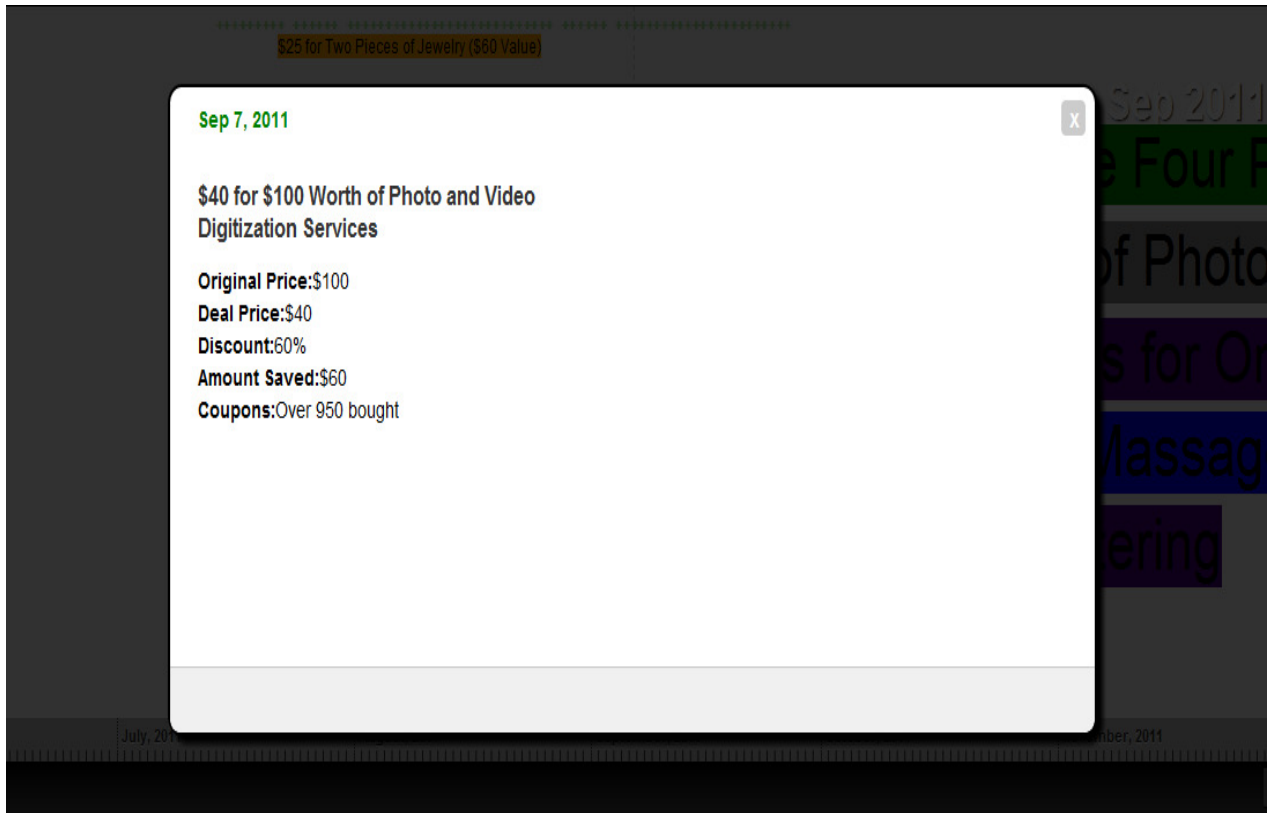
Figure 10 – Pop up window for the deals

Figure 10 shows the deal details. From the information it is easily understandable that the deal is a popular one. More than 950 people bought the deal since the discount is 60%. On top of this pop up window we can see the starting date as well. As far a deal is concerned the above information is the crucial ones. This is the reason why we have restricted the information in the pop up to these details.

3.3 Summarized Visualization

The next step is to provide an overall concise picture of the deals sales across US. The idea is to get a complete picture on deals sales for each state of US. So pie charts are one way of representing this.

3.3.1 Pie Chart

Pie charts are one of the best ways for representing statistical data in business world. The proportion occupied by the arc symbolizes the quantity in represent in the entire scale. One major reason for choosing pie chart is because we have finite number of deals types and hence their representation / analysis will be easier in a pie chart.
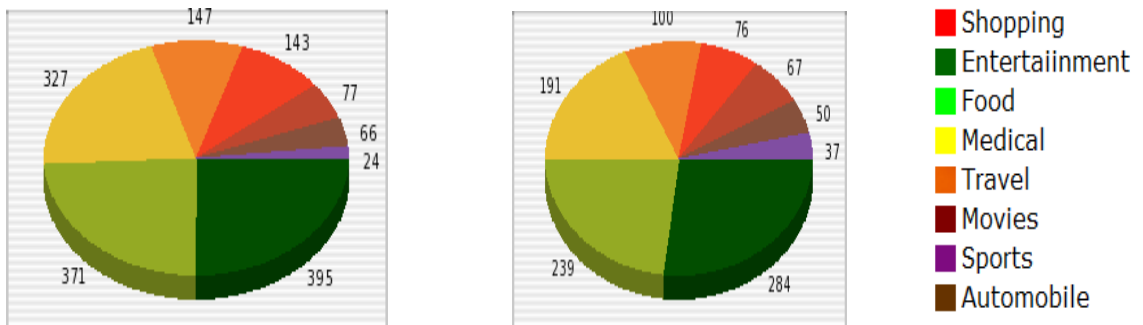

Figure 11 – Large Pie Charts - California and Texas

After careful analysis, we have found that deals related to entertainment like movies, shows, theme park tickets, etc … dominate the chart and more people are interested in buying them.
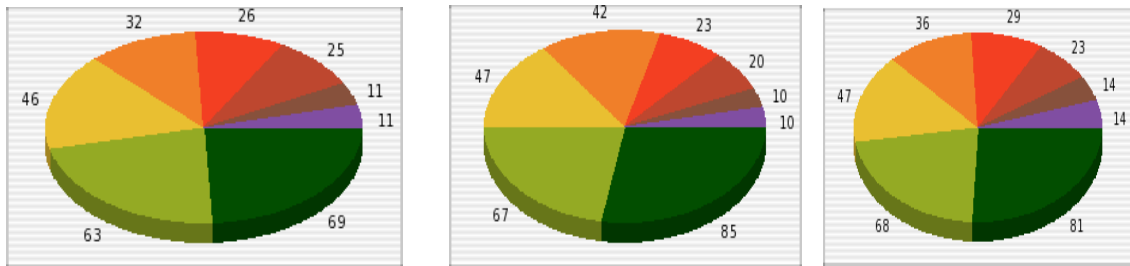
Figure 12 – smaller Pie Charts – Ohio, Oklahoma and Maryland

3.3.2 Visualization

Now to get the complete visualization, we need to incorporate the pie charts on top of their respective states in a US map. We used to help of SVG for achieving this. SVG (Scalable Vector Graphics) is a family of xml based file format for 2D vector graphics. SVG images behave as xml text files. They can be scripted, indexed, searched and compressed. SVG drawings can be animated and made dynamically interactive. Number of event-handlers work on SVG, the commonly used ones being onmouseover and onclick. SVG <g> tag allows bringing a number of other tags (images in our case) together. This grouping helps transformation as a whole easier. The event handlers can be applied to the entire grouping as well as to the individual graphic elements.

```
<g id="usamap">
  <image
     xlink:href="unitedstates.gif"
     y="0.27393264"
     x="7.9117861"
    height="760.0589"
    width="734.17651"
  />
  <image
  xlink:href="images/pennslyvania.png"
  id="pennslyvania"
   y="249.59747"
   x="594.91174"
```

```
   height="51.117649"
   width="61.647057"
  onclick="onclick="window.open('large_pennslyvania.html','mywindow',width=1400,height=800')"
/>
</g>
```

The above piece of code shows part of the SVG code written where the images are interwoven using the <g> tag. The images are placed in the location given by x & y attributes. The length and width of the pie chart are kept proportional to the number of deals sold and type of the deal with the most sales (in our case its mostly entertainment deals). The <image> tag is used mostly in svg than <img> tag. The purpose of both the tags is same but <image> is preferred because we place an image on top of another and this is easily achievable using <image> rather than <img>. They are calculated as modulus of the respective values.

The categories Shopping, Movies, Food, Entertainment, Automobile, Medical, Travel and Sports were based on standard classifications for online deals.

Figure 12 shows a summarized view for all the states. We can clearly see that larger states like California, Florida and Texas have comparatively large number of deals sales and this is because of the number of cities that can be targeted by sellers. Some of the states in central US have no major deals sales that can be represented. Most of the deals website concentrate areas either cities in east coast or west coast where dense population can be found. As a general observation it can be seen that entertainment deals are the most sold ones in many states. In major cities entertainment related deals account for nearly 25% of their total deals sales.
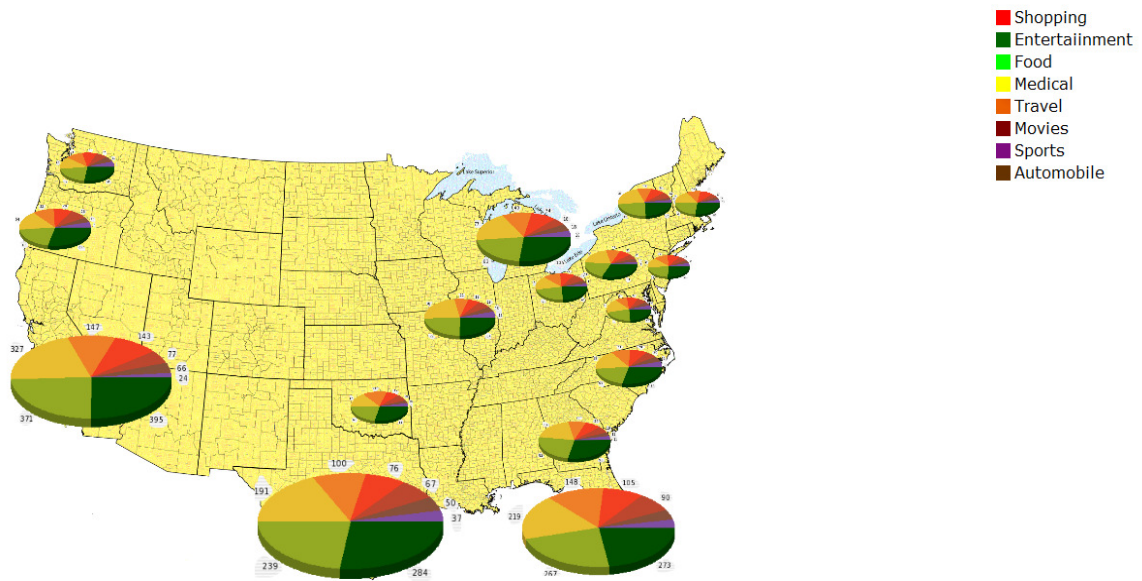
Figure 13 – US map with state wise deals summarized

4. Data Analysis

It has been found that entertainment deals are the most sold in many states. It has been closely followed by Food related and medical services deals. For the state of California, 424 entertainment deals were sold over 3 months from July 2011 to September 2011. This is about 25% of the total deals sold. Deals related to restaurants, food coupons, grocery coupons etc… accounts to about 22% of the total deals. In all, these 2 types of deals account for nearly 47% of the total deals sales. Similarly for Texas, entertainment deals and food deals account for 50% of the deals sales.

Among the food related deals, the coupons from restaurants are the most preferred ones. This clearly shows that people have the psychology to try different cuisines when suitable offers are available. Surprisingly there seems to be an increase in the sales of medical and

health related coupons. Entertainment, food, medical comprises of many items within them. They represent a very large set of widely separated items.

Movies related coupons represent a very narrow category but we can see that they account for 4 – 7% coupon sales in almost all states. So we can clearly say that movie coupons are the single largely dominated category. A standard table can be drawn to explain this. Table 1 shows and compares all the categories of deals for major states.

State	Entertainment	Food	Medical	Shopping	Travel	Auto	Movies	Sports
California	25	22	20	9	9	6	5	4
Texas	27	18	23	10	7	4	6	5
Florida	23	18	22	9	12	3	8	4
New York	24	23	23	9	9	3	7	3
Illinois	23	25	23	6	6	5	6	7
Massachusetts	23	26	17	5	10	4	9	6
Georgia	25	28	19	8	7	3	4	4
Washington DC	21	24	19	6	8	5	9	6

Table 1 – Comparison chart

The above table shows the percentage of various types of deals sales. In all the states, deals related to entertainment, food and medical services accounts for nearly 65% of the total deals sales.

Our idea in this project is come up with visualization so that the above table can be avoided for analysis. People should be able to find the statistics by looking at the visualization rather than read through a table. But for comparison purposes we have planned to use the table format.

5. Conclusion and Future Works

In this project, we have analyzed the historical data and have identified the trend in the sales of online deals. The main idea is to identify how profitable these deals are for both the buyers and sellers. Since the history details are not available for free, the data extraction part is the one that consumed more time and needed more research. We intended to create more interactive data visualization for these deals which would help users identify and understand what they are looking in just a glance or few clicks. The reason for the usage for timeline chart is because representing historical data in a timeline chart will help users look for a particular detail based on a timestamp without much background knowledge about it. The information they are looking for will be available as they move the timeline window either way. Finally providing a nationwide summary representation will help us in getting a much better and bigger picture of the impact these deals have on the respective states.

Regarding the future works, we can make further more changes to this implementation:

- Collect seasonal deals details and identify the seasonal sales
- The deals could be more itemized and more detailed information about each deal can be identified / displayed.
- We have used only the deals of Groupon, we can built a bigger database by collecting data from other major deals selling websites like living social, coupon.com.

Reference

[1] Chang, P., Kim, W., & Yih, W. (2005, April 03). Mining Online Deal Forums for Hot Deals. *IEEE/WIC/ACM International Conference 2005*, 384-390.

[2] Creemers, M., Heijden,H., & Verhagen,T. (2003). Understanding online purchase intentions: contributions from technology and trust perspectives. *European Journal of Information Systems,* 12(1).

[3] Buja, A., McDonald, J., Michalak, J., & Stuetzle, W. (2002, August 6). Interactive data visualization using focusing and linking. *IEEE Conference on Visualization 2002*, 156-163,419.

[4] Fraser, S., & Han, J. (1996, June 2). Data mining techniques. *ACM SIGMOD International conference on Management of data 1996,* 25(2), 545.

[5] Butler, P. (2011, February 28). What $480M of Gross Revenue Looks Like to Groupon. Retrieved from http://www.r-bloggers.com/what-480m-of-gross-revenue-looks-like-to-groupon/

[6] Facebook Graph API. Retrieved March 12, 2012 from http://developers.facebook.com/docs/reference/api/

[7] Scalable Vector Graphics – Wikipedia (n.d.). Retrieved March 13, 2012 from http://en.wikipedia.org/wiki/Scalable_Vector_Graphics

[8] Deals of the day – Wikipedia (n.d.). Retrieved March 24, 2012 from http://en.wikipedia.org/wiki/Deal_of_the_day

[9] Timeglider. Retrieved March 24, 2012 from http://timeglider.com/

[10] Short URL. Retrieved march 24, 2012 from http://hasin.me/2009/05/05/expanding-short-urls-to-original-urls-using-php-and-curl/

[11] Eom, Y., Fortin, D., Hahn, M., Hyun, Y., & Kang, H. (2006, August 29). Effects of perceived behavioral control on the consumer usage intention of e-coupons. Psychology and Marketing, 23(10), 841-864.

[12]Shor, M., & Oliver, R. (2005, November 3). Price discrimination through online couponing: Impact on likelihood of purchase and profitability. Journal of Economic Psychology, 27(3), 423-440.

[13] Deconstructing the Groupon phenomenon. Harvard Business Review July 2011 edition. Retrieved on 25 March, 2012 from http://hbr.org/tablet/0711/vision-statement

[14] Kumar, V., Furuta, R., &Allen, R. (1998). Metadata Visualization for Digital Libraries: Interactive Timeline Editing and Review. Proceedings of the third ACM conference on Digital libraries.

[15] Mena, J. (1999). Data Mining Your Website. Retrieved on March 26, 2012 from http://dl.acm.org/citation.cfm?id=520273