Project 1

\* \_\_\_The Problem\_\_\_: What's the background and scope of the project idea? What problem are you attempting to address or solve? Who may it matter to?

‘All In Melbourne’ is a website offering visitors information on what's on in Melbourne with all the latest events, restaurants, shopping destinations and things to do in the city.

The website offers information on 700 events and 3,000 businesses approximately on daily basis to the public.

The current website is being redevelop and would like to be able to recommend future content to visitors based on their preferences.

A report into the state of the [U.K.’s top 25 e-commerce fashion brands](http://www.cx-score.co.uk/report.html) has revealed that the strongest sites offering product recommendations saw 140% more page views per visit and 10% higher task completion rates. Which is why the most ambitious brands are investing heavily in building, and improving these systems.

Previous market and user research (online polls, heat map analysis and session recordings) and analysis of customer behaviour (Google analytics data) has proven low engagement and retention levels due to irrelevant content being served by the current search engine. The business has worked on improving the search engine and expects the search matches to be improved importantly with the upcoming website revamp.

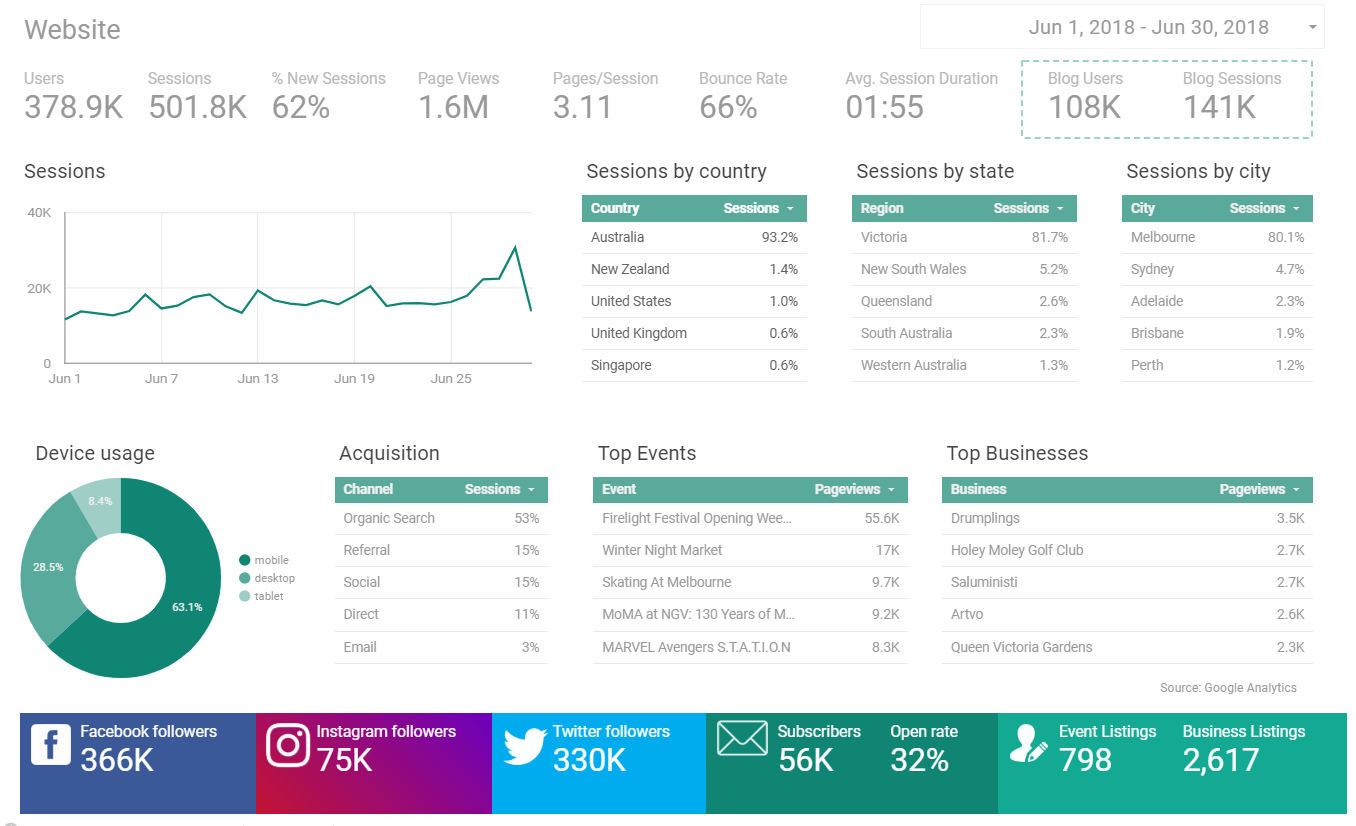
However, the last piece in the puzzle is to serve more relevant/personalised content to the audience by recommending similar content and/or different content to users based on their own preferences. I hope to being able to build an algorithm that can be integrated with the new search engine to improve customer engagement and retention levels on the site.

We would like users to being able to click on a “You might also like” section to display categorized events and/or businesses that might be of interest based on previous behaviour. Users could also check out bundled services under the “While here why not…?” section where related content would be surfaced. Customers are presented with service recommendations based on search terms, customer feedback, geographic location and other online behavioural and product related data.

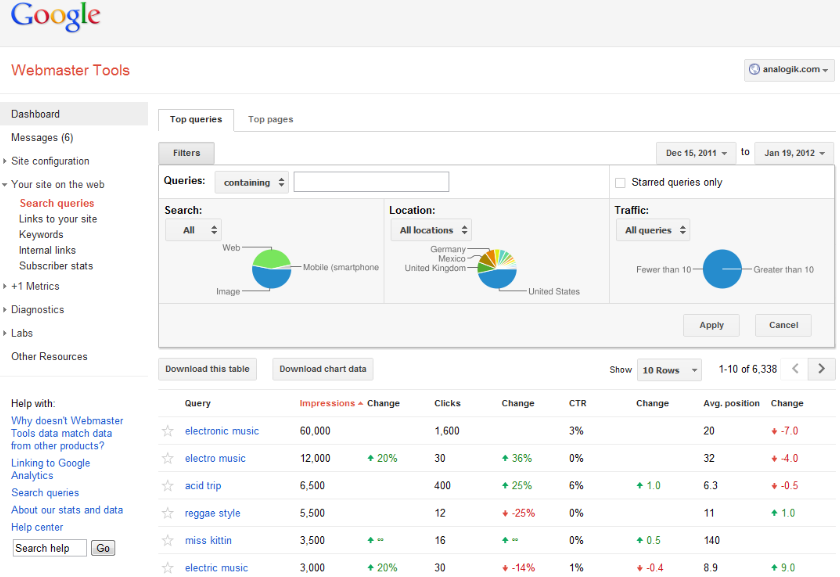
The project is valuable to the end user and therefore to ‘All In Melbourne’ stakeholders (executive board, marketing and digital teams) to improve marketing ROI.

\* \_\_\_Data\_\_\_: What data exists to help solve this problem? Where is it coming from? What does the data look like? What is the observation?

* Website data (Google Analytics)



* SEO data (Google Webmaster)

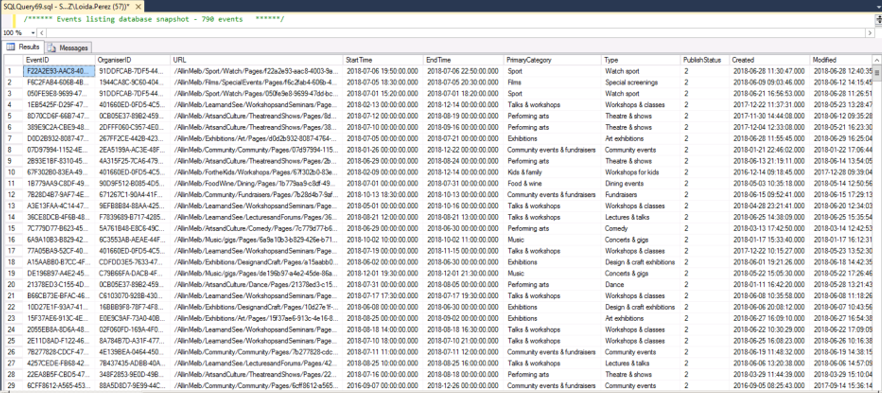


* Social media data (‘All In Melbourne’ Facebook and Twitter groups)

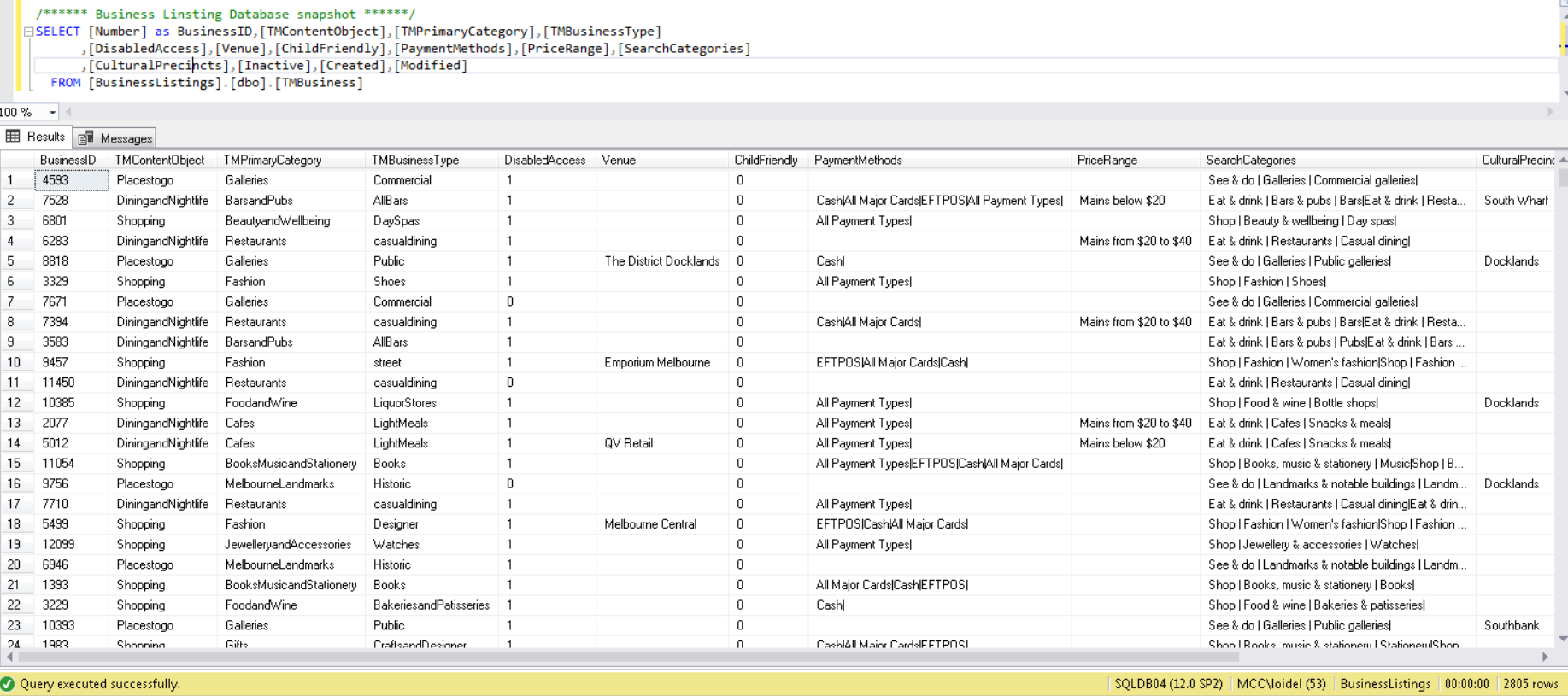




* Events listing database (ELD) (SQL Server database)



* Business listing database (BLD) (SQL Server database)



\* \_\_\_Hypotheses\_\_\_: Given the problem and data you're aware of, what do you believe is the solution? What does success look like?

We would like to explore the association between 1. Customer retention and engagement on the site and 2. Offering more relevant content via recommendations to ‘All In Melbourne’ audience based on the following alternative hypothesis:

*Customers are more engaged and retained with the site when served more relevant content via recommendations*

The null hypothesis is that serving more relevant content thanks to recommendations has no impact in customer engagement and retention.

For the project purposes we define customer engagement as:

* Increase in x minutes in session length
* Increase in x times in session frequency
* Increase in x in organic search rate
* Increase in x minutes in time on page

and customer retention as:

* Increase in x retention rate (as the opposite of bounce rate)

I expect to use CRM, website, search and perhaps social data available from the company. Most of the data is currently available but needs to be retrieved and thoroughly cleaned. I expect data cleansing and manipulation to take the bulk of my time on the project and to become the biggest challenge.

The technique of choice for the search engine recommendation could be multi-variable regression, decision trees, K-nearest neighbour (KNN), support vector machines (SVMs), neural networks and naive Bayes

The success of this project will mean that the relationship between customer engagement and retention and the recommendation of content stated exists.

The project will be documented at my personal repository in GitHub GA account and expect to be delivered on around mid-August when the course is completed. The time allowance foreseen on the project would be two days/week (Saturdays and Sundays)

Project 2

\* \_\_\_The Problem\_\_\_: What's the background and scope of the project idea? What problem are you attempting to address or solve? Who may it matter to?

‘All In Melbourne’ blog currently has low visitation due to its content not being properly surfaced within the website. The blog is treated as an isolated item instead of being integrated with the rest of the website. The website’s redevelopment includes the integration of the blog content into the future site.

However, the company would like to being able to predict the popularity of blog posts in the future based on the current articles so they can determine which content is more interesting to the audience and prioritise it in the new site.

Previous user research via heat maps and session recordings has shown that users are more likely to click and hover on the post titles and categories when at the homepage. There’s also some degree of social media sharing happening across the most successful articles.

The project is valuable to the end user and therefore to ‘All In Melbourne’ stakeholders (executive board, marketing and digital teams) to improve marketing ROI. Being able to predict the most successful articles in the future would improve the company’s content marketing results.

\* \_\_\_Data\_\_\_: What data exists to help solve this problem? Where is it coming from? What does the data look like? What is the observation?

* Website data (Google Analytics) – Same as project 1
* Social media data (‘All In Melbourne’ Facebook and Twitter groups) – Same as project 1

\* \_\_\_Hypotheses\_\_\_: Given the problem and data you're aware of, what do you believe is the solution? What does success look like?

We would like to explore the association between 1. Blog post popularity and 2. Post title, category and social media shares based on the following alternative hypothesis:

*Blog posts popularity is closely associated to amount of shares in social media, post title and category. If the relationship exists and it’s significant we would like to identify the key themes for each of these independent variables.*

The null hypothesis is that the blog posts popularity don’t depend on social media shares, post title and categories.

For the project purposes we define blog post popularity as blog posts with more than 100 pageviews until date of being unpublished.

I expect to use website and social data available from the company. Most of the data is currently available but needs to be retrieved and thoroughly cleaned. I expect data cleansing and manipulation to take the bulk of my time on the project and to become the biggest challenge.

The technique of choice for the could be multi-variable regression.

The success of this project will mean that the relationship between post popularity and shares in social media, post title and category stated exists.

The project will be documented at my personal repository in GitHub GA account and expect to be delivered on around mid-August when the course is completed. The time allowance foreseen on the project would be two days/week (Saturdays and Sundays)

Project 3

\* \_\_\_The Problem\_\_\_: What's the background and scope of the project idea? What problem are you attempting to address or solve? Who may it matter to?

‘All In Melbourne’ would like to improve the overall customer satisfaction and are interested in measuring it on social media, mainly via its Facebook group. At the moment, there’s little visibility on how users feel about the service apart from some polls run in the past.

Previous market research shown that users were unhappy with the content served by the search engine. The issue is currently being improved but nevertheless the company would like to investigate further on customer satisfaction and specially on social media.

The marketing department run a customer segmentation for Facebook users which determine three distinctive segments:

1. Social butterflies (Hyperconnected, device agnostic, socially influential, skew to female users, younger and working)
2. Observers (Passive presence on Facebook, skew to female users, older and busy professionals)
3. Regional occasional (Mostly uses Facebook when visiting the city, living in regional Melbourne, skew to female, mothers, unemployed)

This is why ‘All In Melbourne’ would like to predict the content sentiments of Facebook posts (comments, likes, share) on the ‘All In Melbourne’ group for the previous three segments identified.

The project is valuable to the end user and therefore to ‘All In Melbourne’ stakeholders (executive board, marketing and digital teams) to improve marketing ROI. Being able to predict the sentiment of key target groups on social media would improve the company’s marketing results.

\* \_\_\_Data\_\_\_: What data exists to help solve this problem? Where is it coming from? What does the data look like? What is the observation?

* Website data (Google Analytics) – Same as project 1 and 2
* Social media data (‘All In Melbourne’ Facebook group) – Same as project 1 and 2

\* \_\_\_Hypotheses\_\_\_: Given the problem and data you're aware of, what do you believe is the solution? What does success look like?

We would like to explore the association between 1. Overall customer satisfaction and 2. Content sentiment on Facebook based on the following alternative hypothesis:

*Customers are more satisfied with the site when feeling satisfied (aka positive sentiment) with the content on Facebook. If the relationship exists and it’s significant we would like to identify the sentiment and level of overall satisfaction for each Facebook segment.*

The null hypothesis is that customer satisfaction doesn’t depend on social media sentiment.

For the project purposes we will retrieve comments, likes and shares through Facebook’s API during a 7 days period. I expect to use website, and social data available from the company.

Most of the data is currently available but needs to be retrieved and thoroughly cleaned. I expect data cleansing and manipulation to take the bulk of my time on the project and to become the biggest challenge.

The technique of choice for the project could be natural language processing, classification modelling and machine learning (Random Forest, and Support Vector Classifier) sentiment analysis. The success of this project will mean that the relationship between customer satisfaction and customer sentiment on Facebook exists.

The project will be documented at my personal repository in GitHub GA account and expect to be delivered on around mid-August when the course is completed. The time allowance foreseen on the project would be two days/week (Saturdays and Sundays)