



REPORT 1

DATA ANALYSIS

Prateek Majumder
prateekmaj21@gmail.com

The Data

The dataset contains 12043 transactions for 100 customers who have one bank account each. Transactional period is from 01/08/2018 - 31/10/2018 (92 days duration). The data entries are unique and have consistent formats for analysis. For each record/row, information is complete for majority of columns. Some columns contain missing data (blank or NA cells), which is likely due to the nature of transaction. (i.e. merchants are not involved for Inter Bank transfers or Salary payments) It is also noticed that there is only 91 unique dates in the dataset, suggesting the transaction records for one day are missing (turned out to be 2018-08-16).

The columns are-

```
['status', 'card_present_flag', 'bpay_biller_code', 'account',  
 'currency', 'long_lat', 'txn_description', 'merchant_id',  
 'merchant_code', 'first_name', 'balance', 'date', 'gender', 'age',  
 'merchant_suburb', 'merchant_state', 'extraction', 'amount',  
 'transaction_id', 'country', 'customer_id', 'merchant_long_lat',  
 'movement']
```

- Findings from working with data. For details of the charts and visualizations please refer to the Python NoteBook.
- Total transaction volume over the dates have ranged from 10,000 to 40,000 usually, mean transactions have ranged from 100 AUD to 300 AUD usually.
- Male customers have spent more on their transactions over the dates, as compared to their female customers.
- States have widely distributed transaction over the dates, but New South Wales and Queensland have had a few peaks in transaction amounts.
- State-wise, New South Wales, Victoria and Queensland have had most transaction. The fact that most economic centres and large cities of Australia are in this states also supports the fact.
- Going by total transaction in sub-regions maximum transactions have come from Sydney, Melbourne and Brisbane. These are the largest cities in Australia, and it is logical that largest transactions come from there.
- Going by mean transactions in sub-regions high values have been from Pantapin, Gladesville, Parap and such. But it seems these places have had very few transactions. So we cannot yet determine much about them.
- Total values of sum of customers bank accounts balance have been in the range of 1000000 to 1500000 AUD over the dates.

- The mean and usual account balance of the customers has varied from 10,000 to 20,000 AUD over the times. This does suggest that customers have been saving money, and their assets have grown.
- Male customers have had a higher account balance than female customers, over the dates.
- The mean transaction bar plot shows that the value has remained fairly uniform over the ages, except for a few dips at 24,37 and rises at 35,42 and 52. But, on a similar note, the total transaction has high rises for 22 and 34. It might mean there are many customers in the bank for that age.
- Most of the customers of the bank have made more credit transactions, as compared to debit transactions.
- Working with the map coordinates, it becomes clear that the majority of the transactions are based where the major cities of Australia are.



- Making a wordcloud of customer first name based on Transaction mean amount gives us the names having most spending are- Jonathan, Brian, Melissa, Kenneth, Cindy etc.
- I also do a bit of customer segmentation. I worked more of that on Task 2. Task 2 report will have more detailed analysis.
- The key takeaways we can get is that, most spending is coming from the large cities of Australia and from some key states. The bank should look to expand its operations in these areas.
- Some customers also have tendency to spend more than others. Apart from that, the average bank balance of the customers have increased. It indicates that the bank customers are getting well financially.