ANZ Data Program

- Done by- Prateek Majumder
prateekmaj21@gmail.com

- Approaching the project as a Business Analytics Problem.

Data@ANZ is a contact sport

Discover the role data plays for one of the world's leading financial institutions.

This virtual experience program is **free** for all students

The Problem

A dataset is provided which is designed to simulate realistic transaction behaviors that are observed in ANZ's real transaction data. We have to gather insights from 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).

Getting Started

Data analytics is about solving problems and gaining insights. Let us try to understand the data columns provided to us.

There are various data types for any business problem.
Dividing data types into Primary and Secondary is one way.

Primary data is usually collected at the source. It might be from surveys, interviews, observation and so more. It is usually present for the task at hand.

On the other hand, secondary data is not specifically for the task at hand. It might be past records, sales reports etc. They usually exist beforehand.

Then there are – Nominal, Ordinal, Interval and Ratio data types.

From these, Nominal and Ordinal data types are Non-metric data types and Interval and Ratio are metric data types. Let us discuss more how these are defined.

Nominal- For example, we talk about cars. Porsche 911, Lamborghini Aventador, Nissan Skyline, these give definite identities. Hence, they fall under nominal data types.

Ordinal- Now say for example among the three cars, if we sort by top speed, we say Porsche 911 has a higher top speed than Nissan Skyline. Net we denote them as 1st-Lamborghini, 2nd Porsche and 3rd Nissan. Such will be taken as ordinal. Or say someone says, “I prefer Porsche 911 than Lamborghini Aventador”. That will also be ordinal. It usually shows the relation between multiple entities.

Interval- Now, say someone is told to rate the looks of the 3 cars on a scale of 0 to 10, with 0 being not good looking at all and 10 being highly good looking. Say Porsche is rated 7, Nissan is rated 8 and Lamborghini is rated 9. So such data will be considered interval data.

Ratio- Now if we try to determine the exact top speeds of the cars. Say Lamborghini has top speed of 350 kmph, Nissan has top speed of 300 kmph and Porsche has top speed of 320 kmph. Here we get exact values and can carry out a lot of analysis. Such data will be ratio.

Understanding the data given to us

status- It just states that the transaction is authorized or posted.

card_present_flag- This suggests if the account was done via card or not. 1 if via card and 0 if not via card.

Account- Bank account number, a naming for us and a nominal type data.

Currency- Which currency used, here it was AUD, which is Australian dollar.

long_lat – Longitude and latitude of the place where the transaction was made.

txn_description – The transaction description, that is what type of transaction was it. Nominal data as it gives just the types of transaction.

Merchant id, Merchant code and First Name of transaction also, only give us nominal information.

The balance (indicating the bank) balance does give us some Ratio data. Similarly age and amount(Transaction amount) give us some ratio data, a lot of analysis can be done on such data.

Movement, tells us if it is a credit or debit transaction, ordinal data according to me, and similarly gender gives ordinal data of the person being male or female.

To understand the market of the bank, we must see what the data tells us. Customer analytics will help us understand the financial spending of existing customers. An important part of customer analytics is customer segmentation. Segmentation is the grouping of customers who share certain interests, into homogenous groups.

From a business perspective, it is more efficient and easier to pitch a value proposition to a particular homogenous group. It helps us properly understand customer needs. It might depend on the region, where customers are based, their spending habits, amount of money spent and so on.

Understand the day to day proceedings of the bank from a statistical point of view will give the bank an idea of how shape their future proceedings.