ANZ synthesised transaction dataset
Data Predictive Analysis

By
Sai Vivek Amirishetty

Data@ANZ Virtual Experience Program-(Task-2)
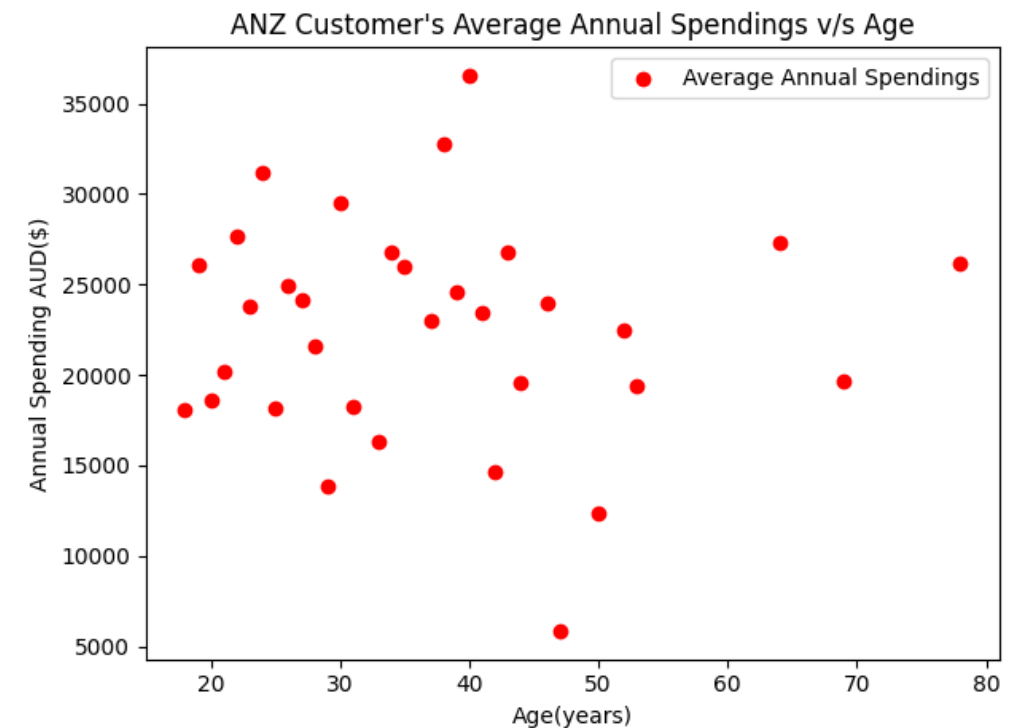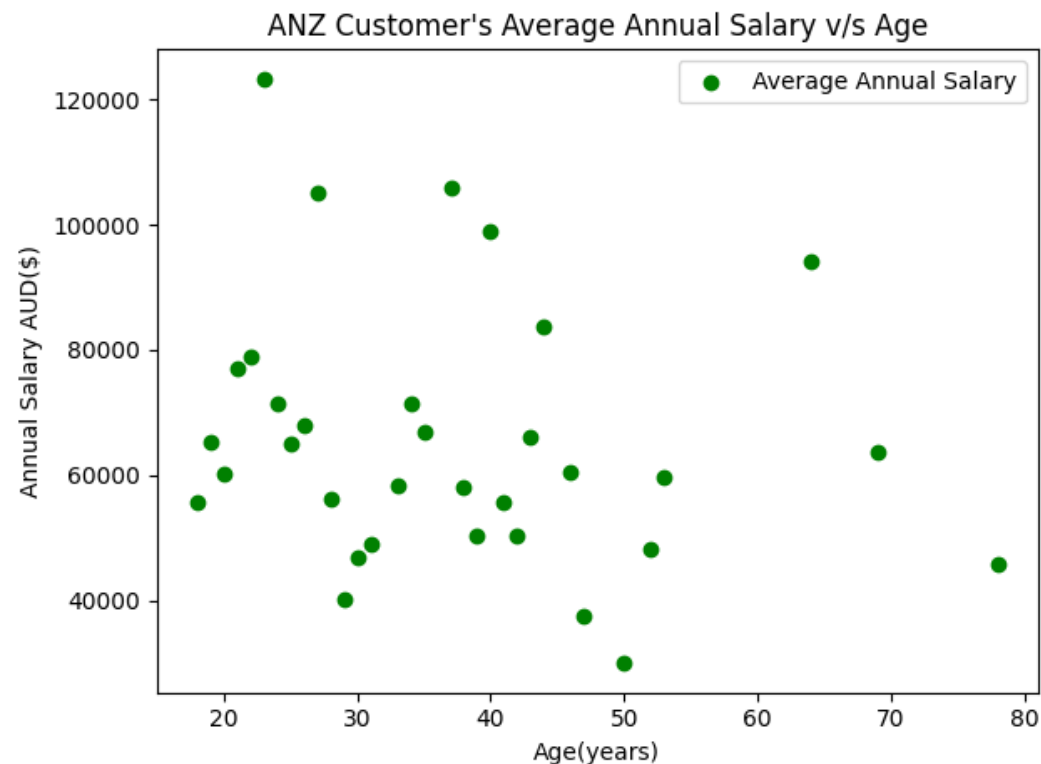
Things Done for the Task

- I have to train two models to predict annual salaries based on the given features, using linear regression and decision trees.

- I used decision tree regressor model to train the categorical data such as age, gender, movement, merchant state, transaction type and linear regression with all the data except annual salary.

```
Mean annual salary by customer:
               card_present_flag  merchant_code      balance  age       amount  annual_salary
customer_id
CUS-1005756958          0.812500            0.0   2275.852055   53   222.862603          50464
CUS-1117979751          0.826923            0.0   9829.929000   21   339.843700         100202
CUS-1140341822          0.815385            0.0   5699.212250   28   212.632500          45996
CUS-1147642491          0.750000            0.0   9032.841186   34   245.600169          88992
CUS-1196156254          0.785276            0.0  22272.433755   34   147.145796         109304

Mean annual spending by customer:
               card_present_flag  merchant_code      balance  age       amount  annual_salary  annual_spending
customer_id
CUS-1005756958          0.812500            0.0   2275.852055   53   222.862603          50464            14611
CUS-1117979751          0.826923            0.0   9829.929000   21   339.843700         100202            35735
CUS-1140341822          0.815385            0.0   5699.212250   28   212.632500          45996            22046
CUS-1147642491          0.750000            0.0   9032.841186   34   245.600169          88992            26931
CUS-1196156254          0.785276            0.0  22272.433755   34   147.145796         109304            34898
```

- This shows us the annual salary and annual spendings of the customer based on the data from three months in the given dataset.

- I came to a conclusion that decision tree does a better job than linear regression in predicting.

Data Visualisation

- This visualisations show us the average annual salaries of each age group and we can see that average salaries increase as we get older but when we are nearing retirement age the average is reduced.
- The Annual Spendings visualisation shows us that younger people spend more but people who are in the age group 40-50 spend more amount and as we are nearing retirement we spend less.
- I also found out that older people tend to use internet for transacting money and younger people use it more.

Thank You!

For giving me this opportunity at ANZ's Data@ANZ virtual experience program

LinkedIn Account: https://www.linkedin.com/in/sai-vivek-amirishetty-463a99b7/