# Data@ANZ Program

Predictive Analytics

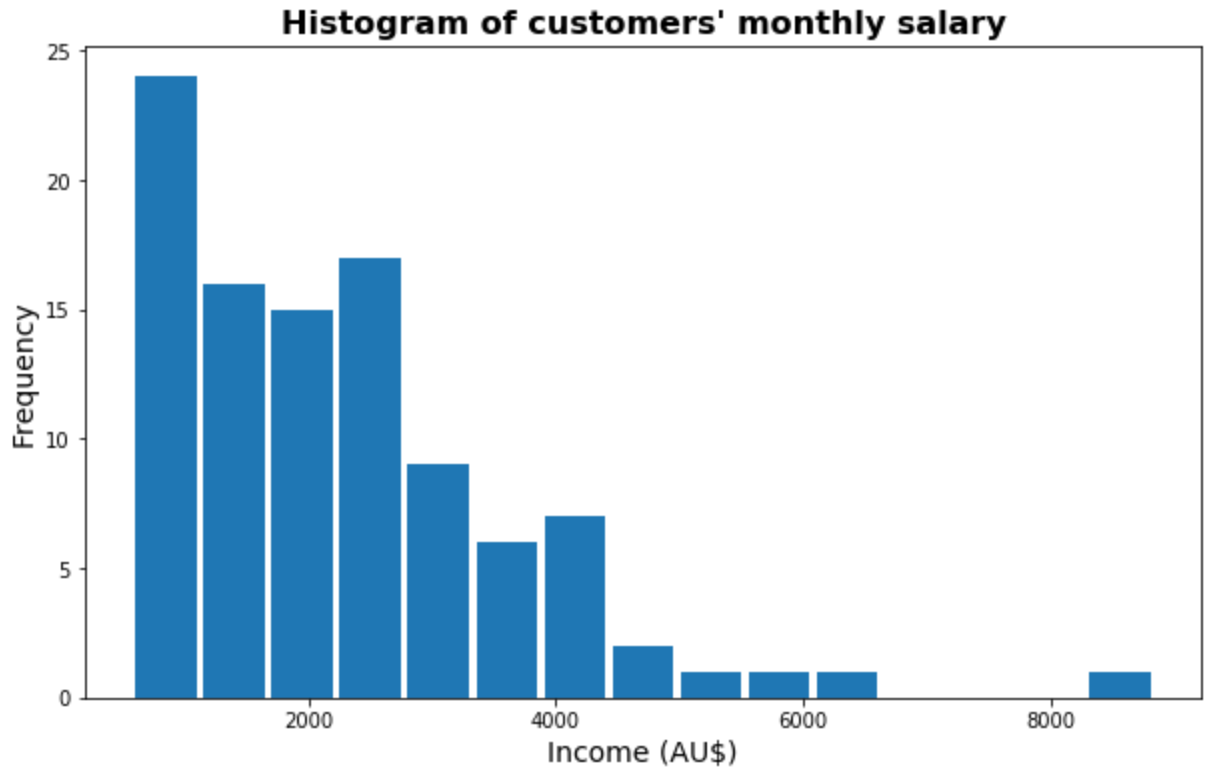
Julio Pimentel

Email: julio.pimentela@gmail.com

Linkedin: https://www.linkedin.com/in/juliopimentel/

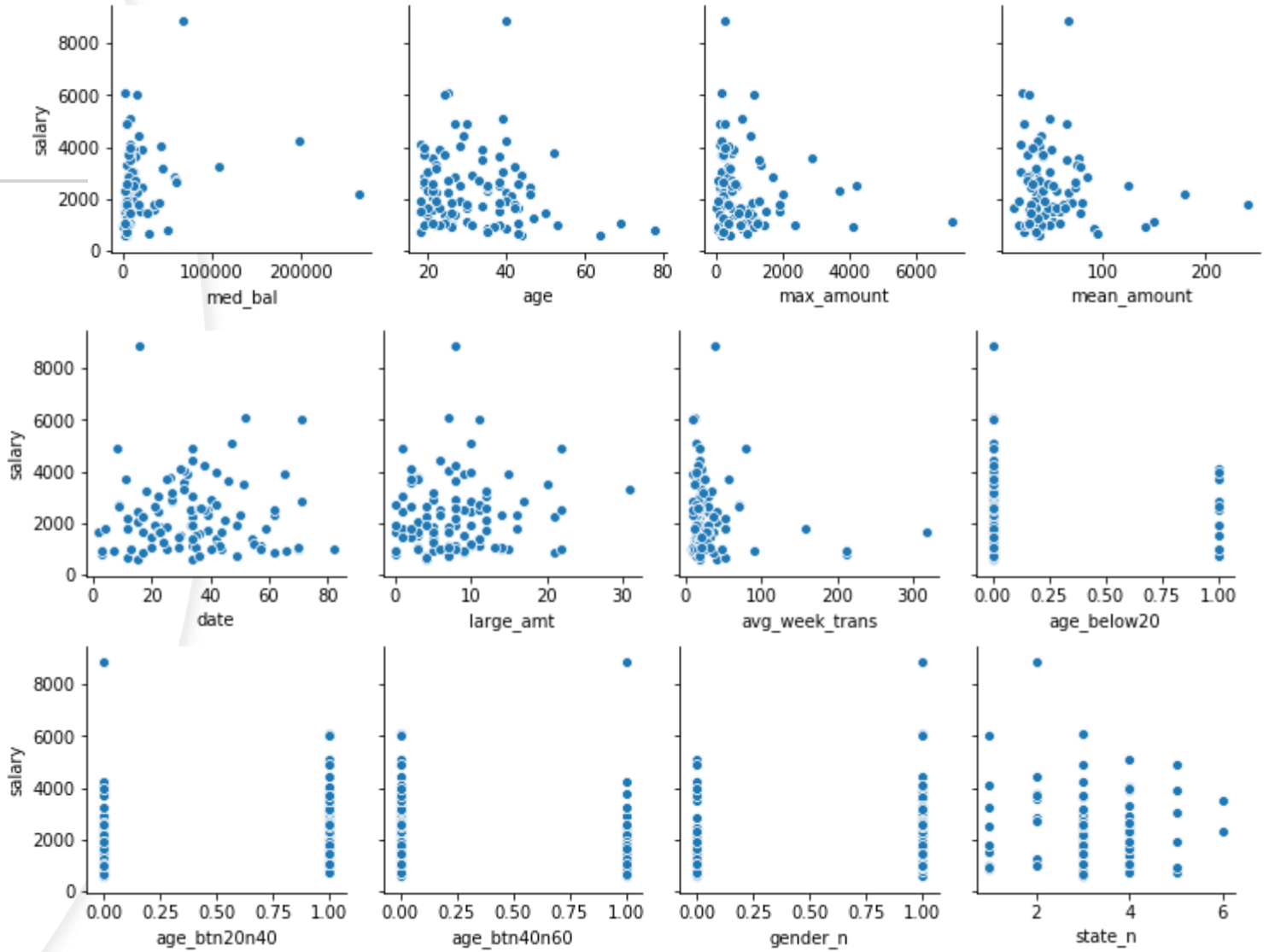
# Insights – monthly salary

- The customer's monthly salary was calculated by filtering the attribute “PAY/SALARY” from the columns “txn_description”.
- It was assumed that the monthly salary did not change in the interval of the dataset.
- 75% of the customers' salary is between AU$576 and AU$2,886.



# Insights – variables

- The original variables of the dataset “age”, and “state” were selected.
- New attributes were created from the original variables:
  - (1) balance’s median, (2) amount’s maximum, (3) amount’s mean, (4) number of dates with a transaction, (5) number of the large transaction amount, (6) average week transactions, (7) age intervals (below 20, between 20 and 40, between 40 and 60).
- Each variable was plotted against the customers’ monthly salary.



# Insights – variables

- The selected variables do not have a significant correlation above 0.50, which is not better than making a guess.
- The first linear regression model using all variables has an $R^2$ of 0.155 and MSE of 1,634,256.
- Second model using the variables “age”, “amount’s maximum”, “average week transactions”, “gender” and “balance’s medium” has a $R^2$ of 0.129 MSE of 1,684,761.