Data@ANZ Program
Exploratory Data Analysis

Julio Pimentel

Email: julio.pimentela@gmail.com
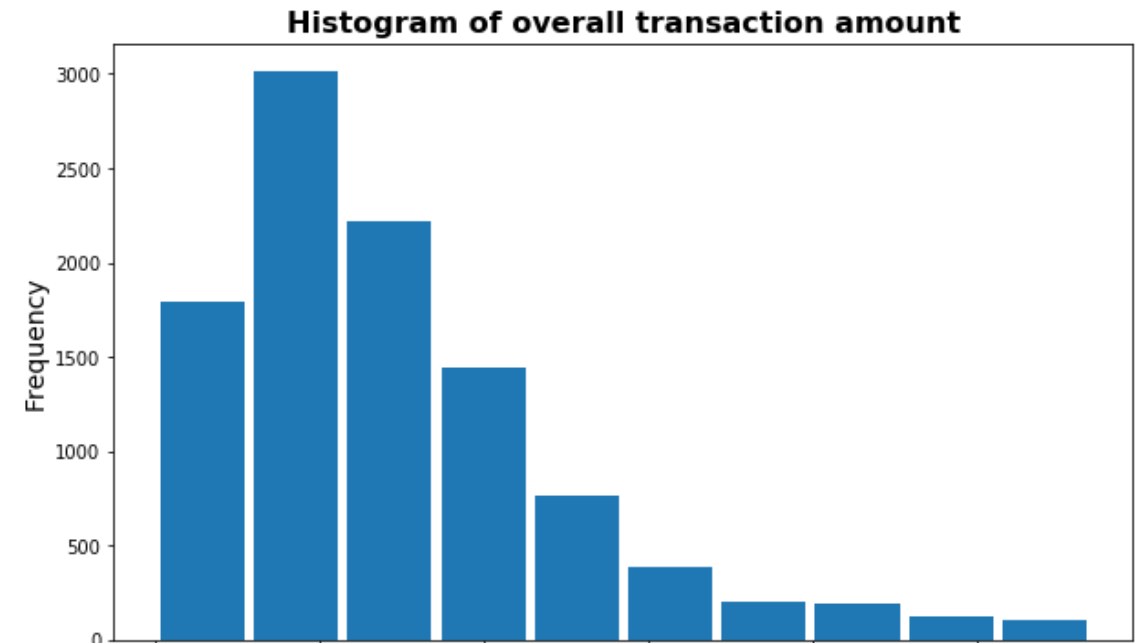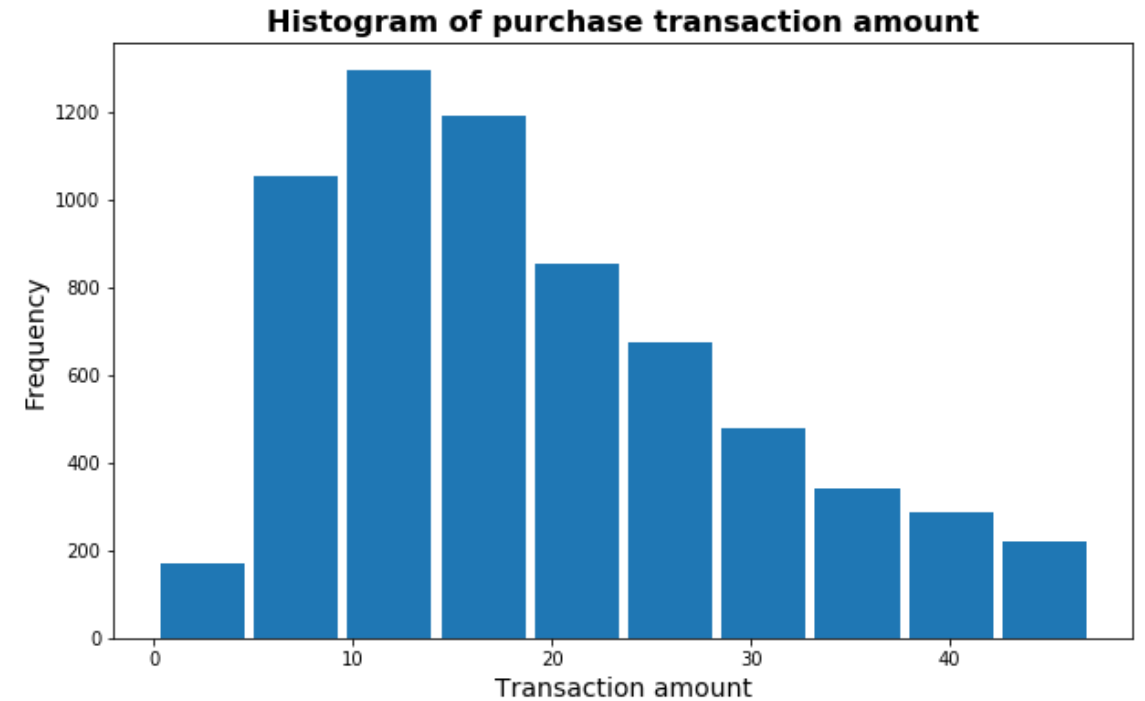
Linkedin: https://www.linkedin.com/in/juliopimentel/

Exploratory data analysis

- The dataset has 12,043 rows.

- Transaction period is from 01/08/2018 to 31/10/2018 with a missing value in 16/08/2018.

- The customers are in Australia except for the customer ID CUS-1617121891.

Insights - transactions

- The purchase transaction values represented 7,717 values of the 12,043 values from the dataset.

- The purchase transaction frequency is concentrated between AU$5 and AU$30.

- The purchase transaction values have 1,158 outliers (above 85% of the sub-data distribution) between AU$47 and AU$7,081.

- The overall transaction frequency is concentrated between AU$0 and AU$40.

- The overall transaction values have 1,807 outliers (above 85% of the-sub data distribution) between AU$114 and AU$8,835.

Histogram of purchase transaction amount

Histogram of overall transaction amount

Insights - spending

- The transaction volume was analyzed by the hour and weekday.

- Wednesday, Thursday, and Friday are the days with more transaction volume.

- Monday and Tuesday are the days with less transaction volume.

- The hours 9:00, 11:00, 17:00 are the highest peak of transaction volume.

- Between 0:00 and 8:00 is the lowest peak of transaction volume.

Average transaction volume by hour

Average transaction volume by weekday