What Was Found?
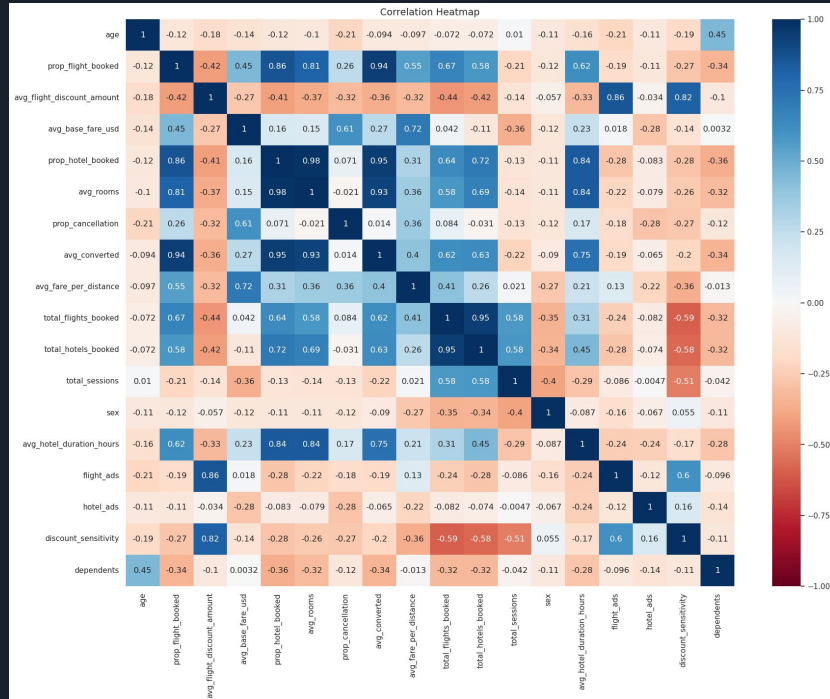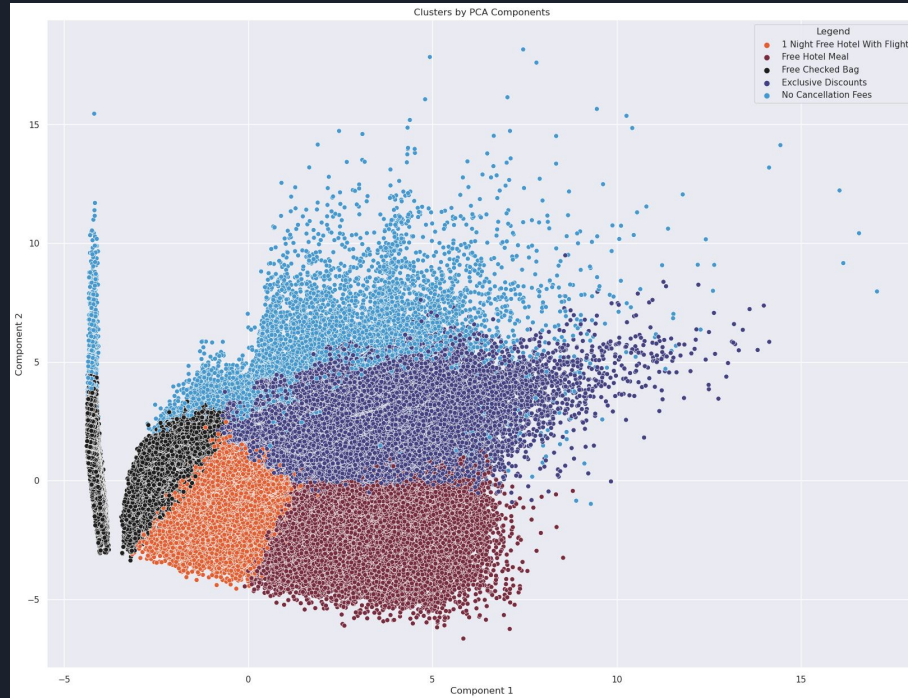
…

Segments.

Overview

Segments Were Visible Upon Using Metrics That Had High Correlation Values.

Correlation Heatmap

Validations Were Possible Using K-Means

The Segments Are Shown Below With Colors And A Legend Indicating Which Is Which

How?

- ○ Correlating The Variables Showed That Many Variables Were Essentially Duplicates Of Each Other, Oftentimes It Was The Hotel/Flight Variables.
- ○ When The Correlation Was High Enough Between Enough Variables It Would Be Left Out.
- ○ Found That The Typical Variables One Might Consider Were Not Of High Importance When It Comes To Travel Tide Customers; sex, marriage, etc.

Actual Findings

- Free Checked Bag: 648,745
- 1 Night Free Hotel With Flight: 228,133
- Free Hotel Meal: 84,294
- No Cancellation Fees: 32,311
- Exclusive Discounts: 27,443

Appendix

- .py files are located in the python folder
- Jupyter notebook .ipynb file is located in the python folder
- User Ids with Attached Segment and Variables is located in the root folder
- Queries are located in the python folder in queries.py as they were used in the jupyter notebook.
- Query1 is the only query used in the finished notebook.