My Submission: Mogbo Emmanuel Ojuba

Banking Customer Clustering and Visualization

Step-by-step methodology

1. Data clean-up
    - Encoding of categorical variables
    - Treatment of null values
2. Data Exploration and Feature Selection
    - Choice of pertinent features
    - Creation of new features
3. Standardization of the variables using sklearn (shifting the features to zero mean and unit variance)
4. Generation of clusters using K-means
5. Inspection of clusters using principal component analysis and scatterplot visualization
6. Description of the clusters and creation of customer personas

Variable Selection and Engineering

Categorical Variable Encoding:

Age, Income and Transaction frequency categorical variables were replaced with a representative numeric variable. For instance, income bracket of “N50,000 to N100,000” was replaced with numerical variable N75,000. Transaction frequency of “At least once a day” was replaced with 30, an estimate of the count of transactions per month. Similarly, “Never” was replaced with 0, “Rarely” with 0.5, etc

Null Values:

Features that had null values for ¼ or more of the records were dropped before they were used in the analysis. The exception to this is in the devices used for online banking sub-section, where a null value under a device (eg ATM) was taken as False i.e the user does not use this device.

Creation/combination of Variables:

Certain variables were created to aid the analysis:

- **Number of Devices Used per person in online banking (dev_per_person**): This variable was created by summing up all the Yes’s in the device used for online banking sub-section.
- **Customer Experience Scores (cust_care_score, complaints_score, branches_score, etc**): These scores were obtained by multiplying the importance rating by the satisfaction rating for each criterion. The products were then summed and grouped under the various sub-sections: customer care, complaints, branches, etc

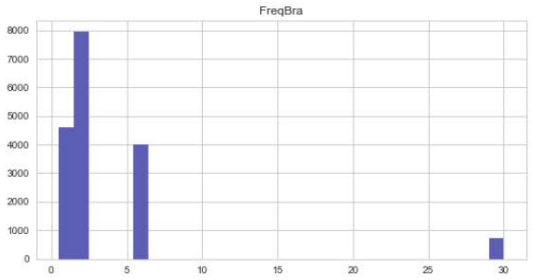
Variable Selection and Engineering

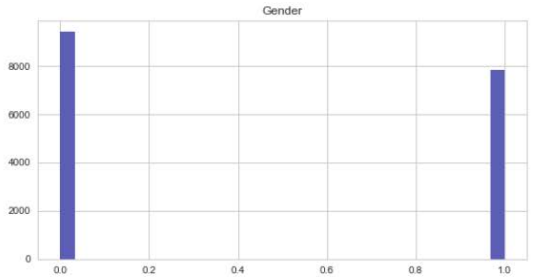
Sub-sections of Survey Not Used in Current Analysis:

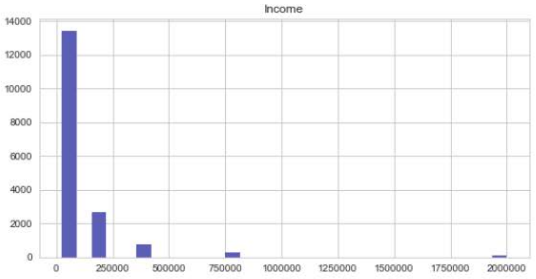
Some sub-sections were not used but might be incorporated into any future analyses:

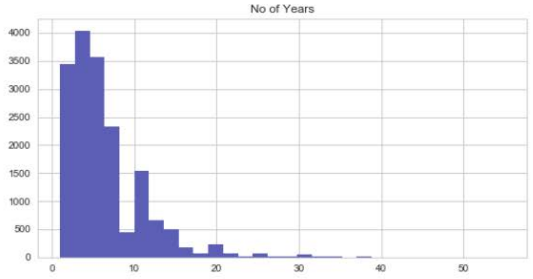
- Possible reasons for maintaining banking relationship
- Reasons for most frequently used bank's ATMs
- Banking activities presently carried out with different banking channels
- Banking activities you would like to carry out with different channels in future
- Payments settled with cash
- Payments settled with non-cash methods
- Payments you will like to migrate to non-cash methods
- Loyalty ratings
- Switching banks

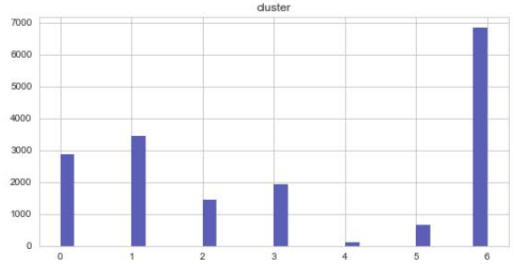
Histograms of Features used in K-Means Run

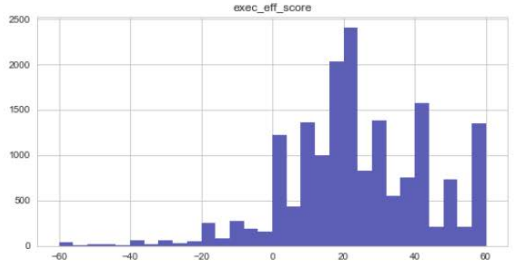


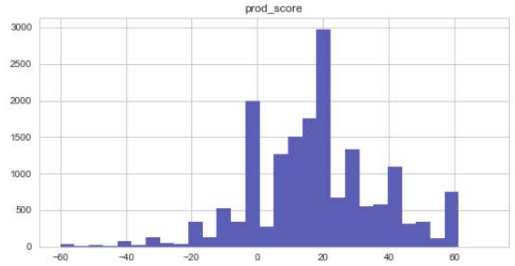


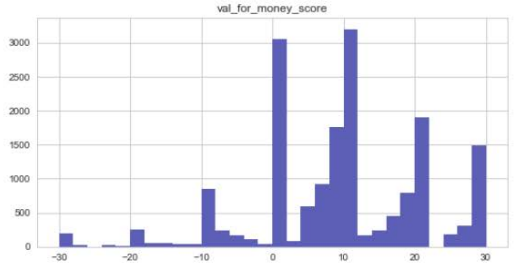


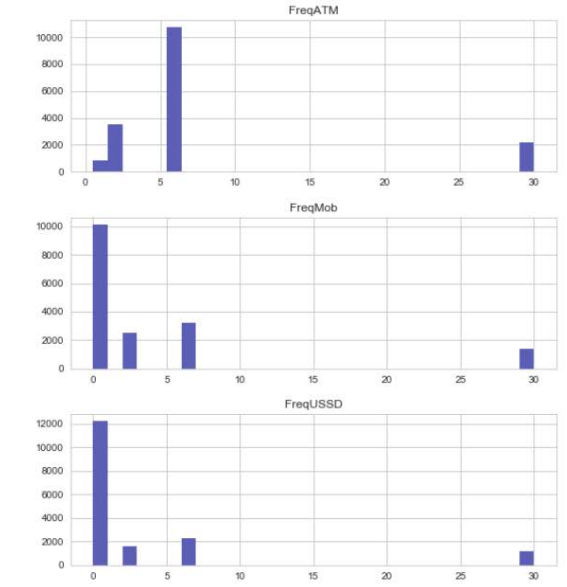


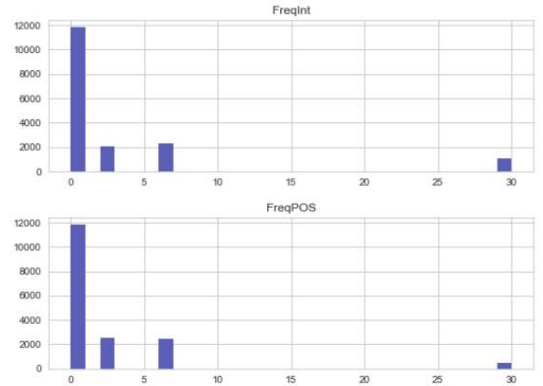


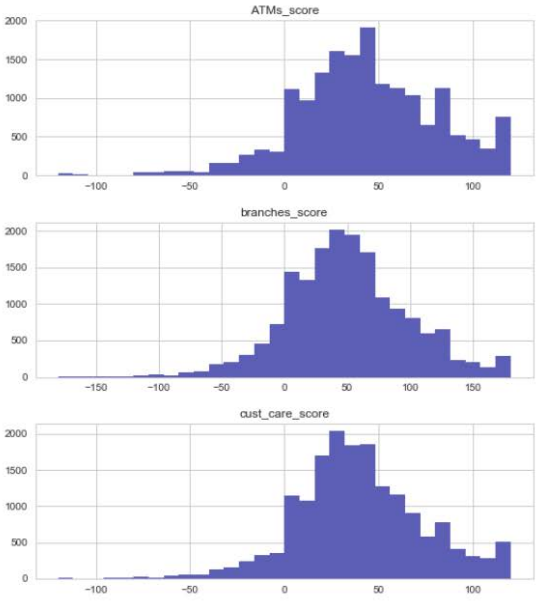


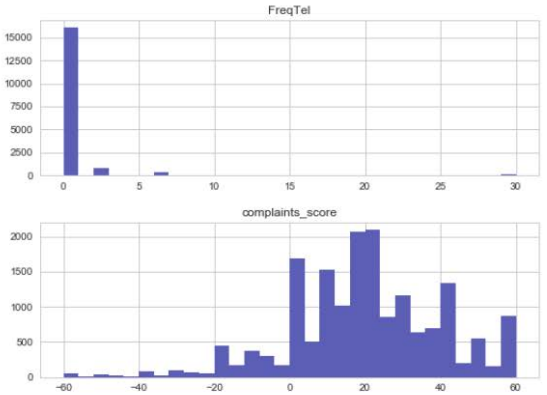












Customer Personas

The Satisfied Young 17%

Age: 31 years
Income 90,000
Loyallty: 5.5 yrs
Branch: 2 – 3 times a month
ATM: twice a week
Mobile, internet: 2 – 4 times a month
Very satisfied customers
Most popular banks: C (20%), B (11%), H (10%)

The Ultra - Transaction Happy 8%

Average age: 30 years
Average income: 130,000
Loyallty: 5.4 yrs
ATM: four times a week
Mobile Internet: 6 times a week
Mildly positive satisfaction levels
Most popular banks: B (24%), A (13%), C(9%)

Aged, loyal, affluent 11%

Age 36 – 60 years
Average Income: 260,000
Loyallty: 14.1 years
ATM: 1.5 times a week
Mobile, Internet usage: 1 – 2 times a month
Moderately positive satisfaction levels
Most popular banks: C (20%), A (11%), H (11%)

Affluent, young, branch visitors 3.7%

Average age: 35 years;
Average income: N230,000
Loyalty: 6.5 years
ATM: 3 times a week
Branch: Daily
Int, Mob Usage: 1 - 2 times a week
Satisfaction: moderately positive
Most popular banks: C (16%), B (14%), H (11%)

The Slightly Dissastifed 20%

Average age: 30 years
Income: 70,000
Loyallty: 5.4 years
Branch: 2 – 3 times a month
ATM: twice a week
Mobile, internet: 1 – 2 times a month
Ambivalent satisfaction levels
Most popular banks: B (15%), C (13%), A (11%)

Young, affluent, transaction happy 0.5%

Average age: 32 years;
Average income: N190,000
Loyalty: 5 years
ATM: 4 times a week
Branch: 2.5 times a week
POS, Int, Mob: 3.3 times a week
Satisfaction: moderately positive
Most popular banks: C (18%), B (13%), D (10%)

Young, broke, few transactions 40%

Average age: 29 years
Income: 60,000
Loyalty: 4.7 years
Branch: twice a month
ATM: 1.5 times a week
Int, Mob: 1 – 2 times a month
Moderately positive satisfaction
Most popoular banks: B (17%), C(13%), A (10%)

Dollar Card Product Introduction (KPMG Question)

Which two clusters will you select for the introduction of a new dollar card product?

- I will select two clusters:
  - Young, affluent, transaction happy: This group is young, has high frequency of mobile/internet transaction and fairly wealthy. They are a good target for dollar product as they have the money to pay for foreign products and are tech savvy.
  - Affluent, loyal, aged: This group has the money and some level of comfort with using mobile/internet for transactions (at least monthly). They show high loyalty to the bank. Clever marketing/education might be needed to convince them to use their card more for foreign transactions.

Further Work

- Use a percentage rating for the customer experience scores to put all the ratings on the same scale
- Investigate the failed t-SNE plot as the first few principal components are not strong enough for 2-D visualizations
- Include *Banking Activities Presently Carried out with Different Banking Channels* sub-section as part of clustering analysis. There might be some trends there
- Examination of each of the clusters for variation, and outliers if any
- Investigate Affluent, young, frequent branch visitors cluster (3.7%) – their frequency of bank branch visitation is daily which is quite unusual