DST Banking Case Study

Sudip Bairagi

Problem Statements

1. The Loans Division of ABC Bank want to know the accounts who are likely to default in repaying the loans when the contract ends.

2. Identify whether any demographic factors is showing strong association with loan defaulters.
3. Identify if there is any relation between the type of credit card issued to a disponentwith his average credit balance maintained in transaction account.
4. Visual dashboard to show -
    Distribution of client age as of their first transaction date
    Distribution of client age by gender from their first transaction date
    Distribution of clients by districts and average salary
    Distribution of clients by average transaction amount
    Any other interesting visual representation of the findings

Technology Used

Python
Pandas , Seaborn and sklearn

IDE
Anaconda Distribution , jupyter and spyder

Database
MS Access : As the data set are in TXT format and Tableu has good integration with Access TXT
Excel : Used it for a simple date formatting calculation on Transaction table

Dashboard and Visual Report
Tableau Online
Dashboard link : https://eu-west-1a.online.tableau.com/#/site/sudipbairagi/views/BankCaseStudy/Dashboard1

Problem 1 : Predict Loan Defaulter

Logical Steps and code snippet

- Import all TXT data into MS Access and Table
- Write SQL Query to join Loan and District table and generate out in CSV : LoanToDemographDistinctMap.xls
- Panda Code : BankingDefaulterPrediction.py

Algorithm used: **K Nearest Neighbour (KNN)**

This algorithm is selected to classify A,B,C and D type Loan Payers with District demographic data.

df = pd.read_excel('LoanToDemographDistinctMap.xls')

Data Cleansing : To replace NaN with 0 and Scale the data using a StandardScaler and remove non-numeric column from DataFrame.

df.fillna(value=0, inplace=True)
scaled_DemoGraphNumeric = scaler.transform(DemoGraphNumeric)
*scaled_DemoGraphNumeric_feature = pd.DataFrame(scaled_DemoGraphNumeric,columns=DemoGraphNumeric.columns)*

Prediction Model : KNN with K=10
knn = KNeighborsClassifier(n_neighbors=10)
knn.fit(scaled_DemoGraphNumeric_feature,df['status'])

Find Best Fit K value

```
error_rate = []
for i in range(1,60):
    knn = KNeighborsClassifier(n_neighbors=i)
    knn.fit(scaled_DemoGraphNumeric_feature,df['status'])
    pred_i = knn.predict(X_test)
    error_rate.append(np.mean(pred_i != y_test))
plt.figure(figsize=(10,6))
plt.plot(range(1,60),error_rate,color='blue', linestyle='dashed', marker='o',
    markerfacecolor='red', markersize=10)
```


Error Rate vs. K Value
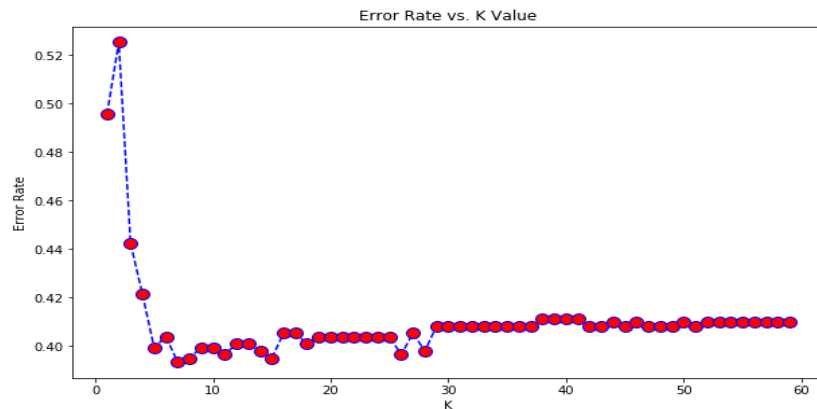
Problem 1 : Classification Report and Confusion Matrix

Confusion Matrix :

```
[[ 44   0 158   0]
 [  6   0  24   0]
 [ 37   0 362   0]
 [  5   0  40   0]]
```

Classification Report:

	precision	recall	f1-score	support
A	0.48	0.22	0.30	202
B	0.00	0.00	0.00	30
C	0.62	0.91	0.74	399
D	0.00	0.00	0.00	45
avg / total	0.51	0.60	0.52	676

Result :

- With optimal value K=10 , the Classification report could not determine the bad defaulter Type B .
- The reason may be because of number of records available for Defaulter against good loan payer .

Problem Card to Avg Loan Balance

Logical Steps and code snippet

- Import all TXT data into MS Access and Table
- Write SQL Query to find the Last Balance from the Transaction Table : AvgBalancePerAccountQuery

Report **Tableau Online link** : **https://eu-west-1a.online.tableau.com/#/site/sudipbairagi/views/BankCaseStudy/BalanceMaintainedPerCardType?:iid=1**
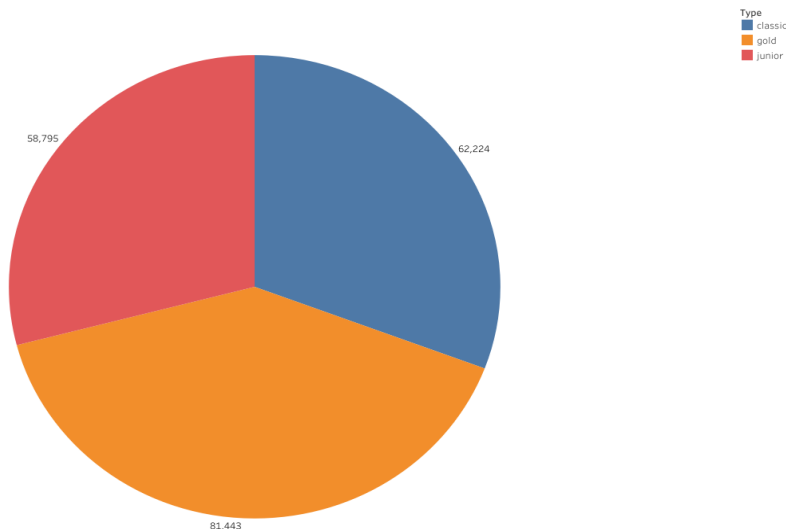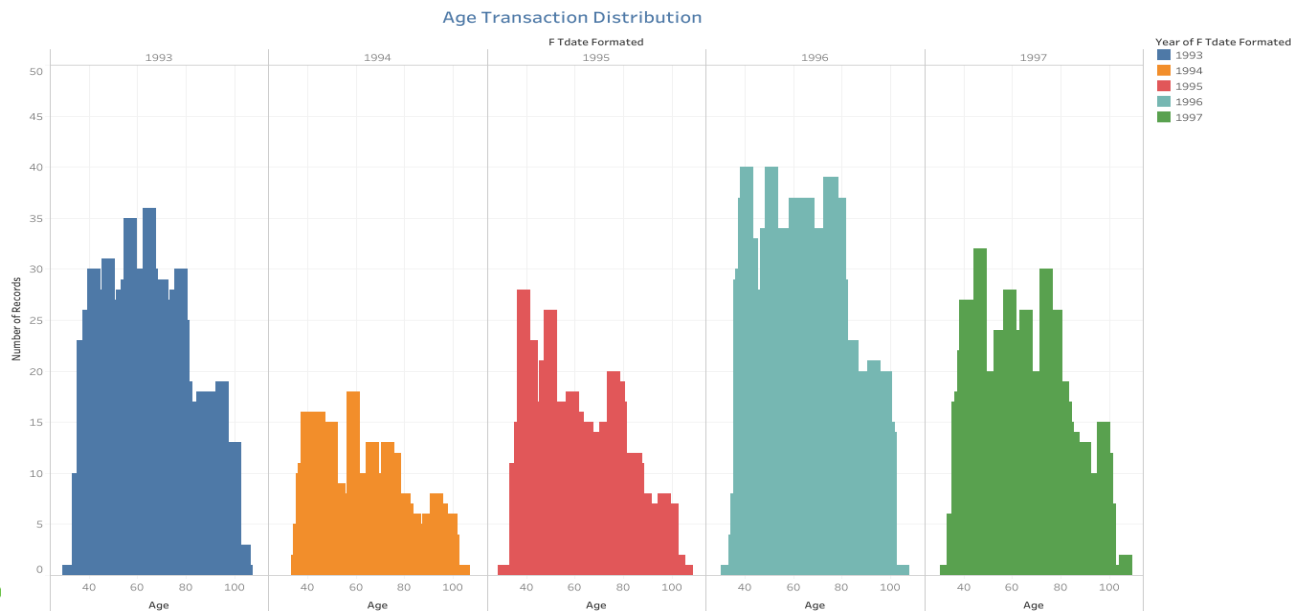
Avg Balance Maintained Per Card Type

The Report in Tableau is interactive one.
Putting cursor on the report shows more data and can be drilled down .

Client age as distribution of their first transaction date

Logical Steps and code snippet

- Import all TXT data into MS Access and Table
- Write SQL Query to find the Age of each Client : **ClientBirthDateResult**
- The Date Number fields is also need to decode to find Male and Female (50+M)
- **Tableau Calculated Fields** are created to calculate the Age using Tableau in-build functions .

Tableau Online link : https://eu-west-1a.online.tableau.com/#/site/sudipbairagi/views/BankCaseStudy/AgeTransactionDistribution?:iid=1
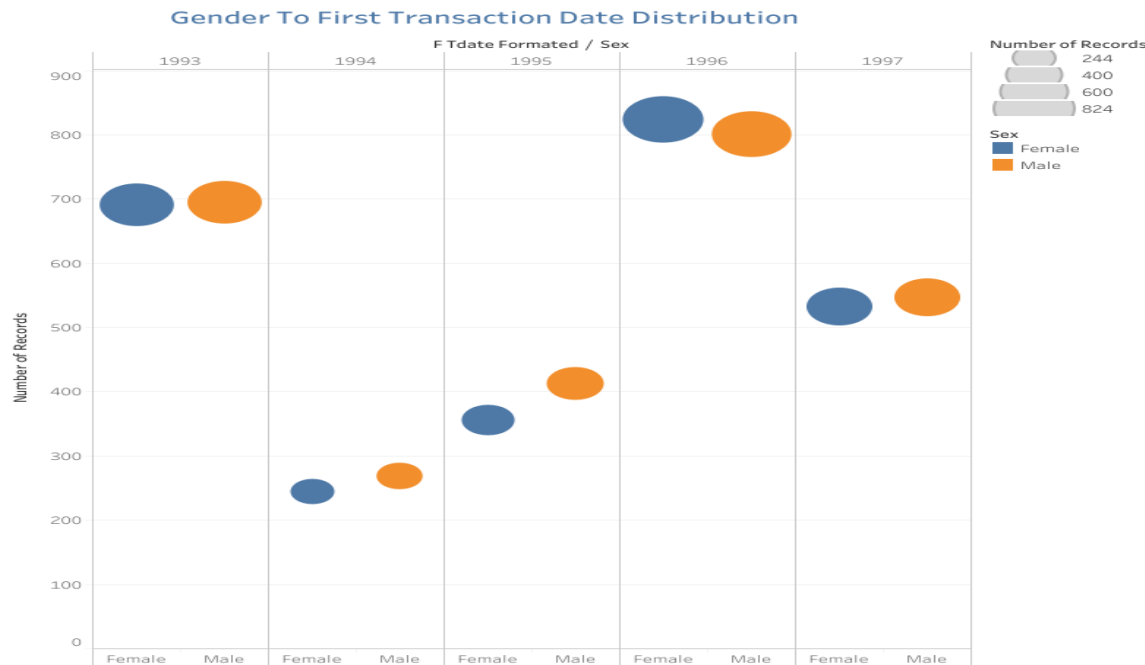

Age Transaction Distribution

The Report in Tableau is interactive one.
Putting cursor on the report shows more data and can be drilled down .
In 1994 it is least amount .
In 1996 it is the Highest amount .
40-50 Years age is having highest First Transaction Counts.

Client Gender as distribution of their first transaction date

Logical Steps and code snippet

- Import all TXT data into MS Access and Table
- Write SQL Query to find the Age of each Client : **ClientBirthDateResult**
- **Write MS Function to calculate Gender from Birth Number : ClientBirthDate Module -> clientSex**
- The Date Number fields is also need to decode to find Male and Female (50+M)

Tableau Online link : https://eu-west-1a.online.tableau.com/#/site/sudipbairagi/views/BankCaseStudy/GenderFirstTransactionDistribution?:iid=1

Gender To First Transaction Date Distribution

The Report in Tableau is interactive one.
Putting cursor on the report shows more data and can be drilled down .
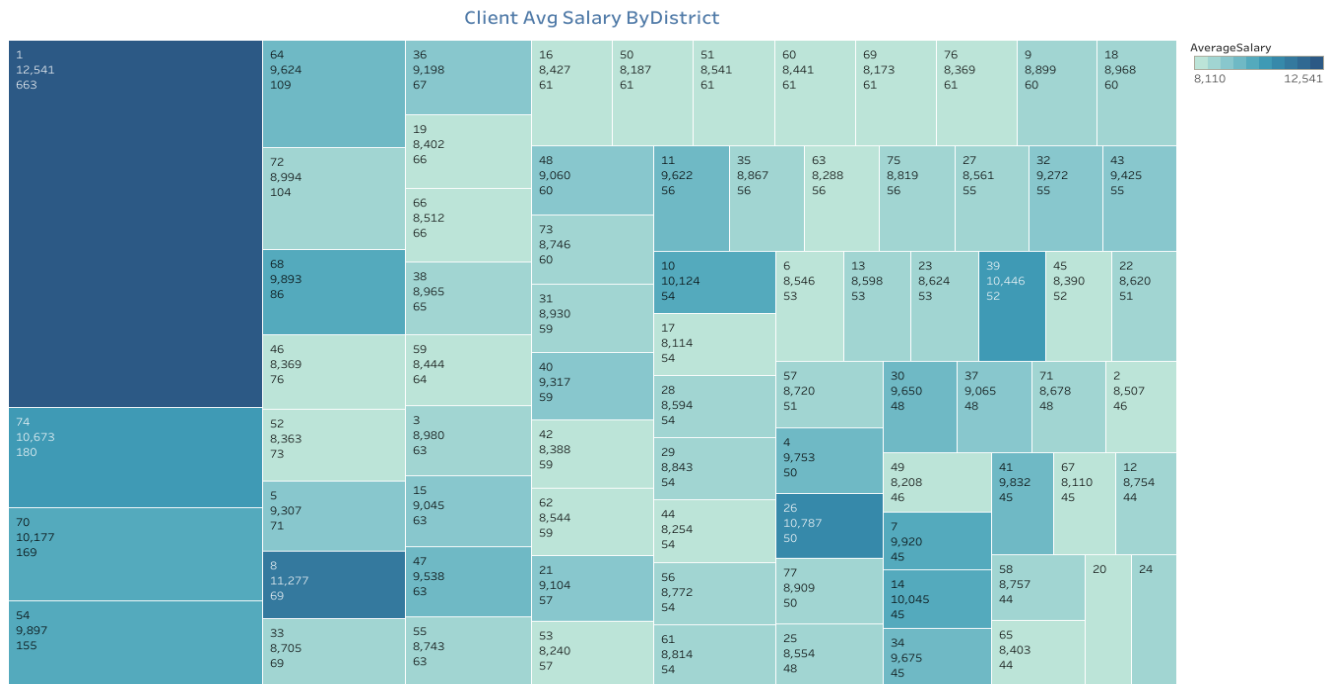In 1996 Female had more first transaction than Male.

Distribution of clients by districts and average salary

Logical Steps and code snippet

- Import all TXT data into MS Access and Table
- Write SQL Query to find the join District

Tableau Online link : https://eu-west-1a.online.tableau.com/#/site/sudipbairagi/views/BankCaseStudy/ClientAvgSalaryByDistrct?:iid=1
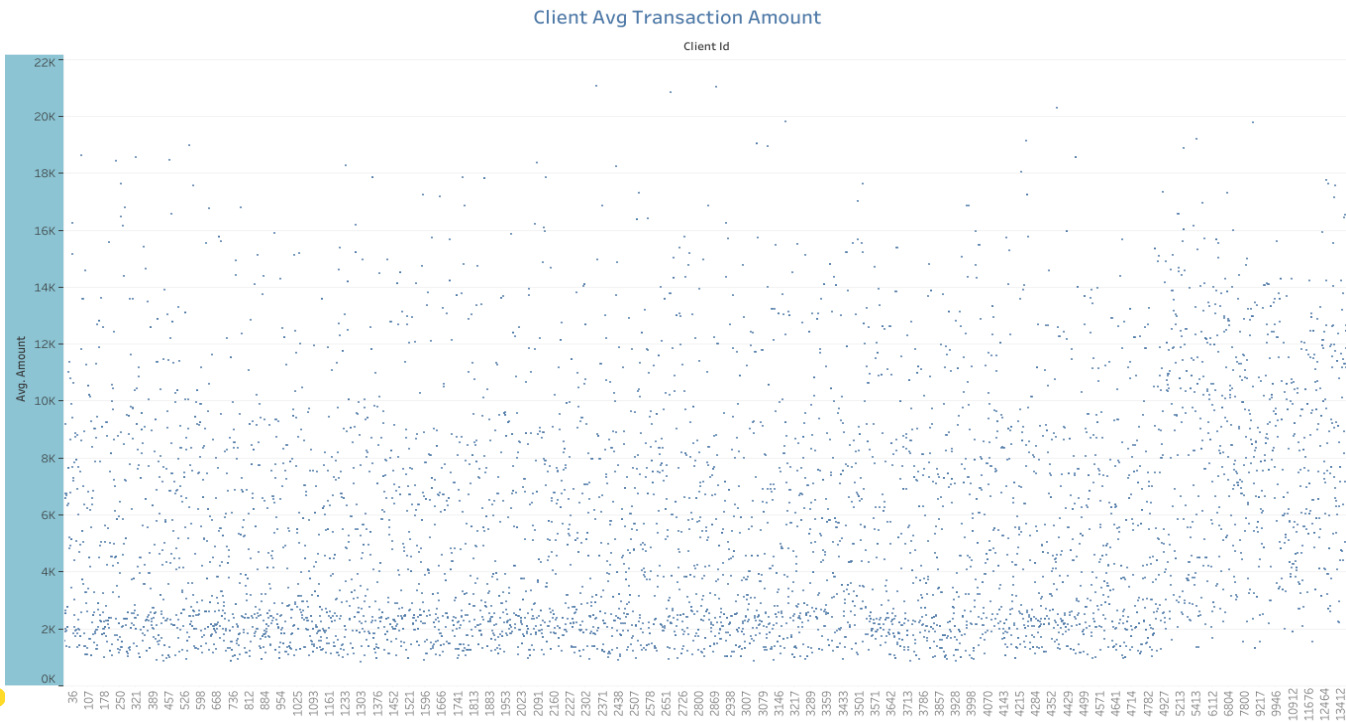

Client Avg Salary ByDistrict

The Report in Tableau is interactive one.
Putting cursor on the report shows more data and can be drilled down .
District 1 has most of the Avg Salary and dense.

Distribution of clients by average transaction amount

Logical Steps and code snippet

- Import all TXT data into MS Access and Table
- Write SQL Query to find the join Transaction with Disp and Client

Tableau Online link : https://eu-west-1a.online.tableau.com/#/site/sudipbairagi/views/BankCaseStudy/ClientAvgTrnAmount?:iid=1

The Report in Tableau is interactive one.
Putting cursor on the report shows more data and can be drilled down .
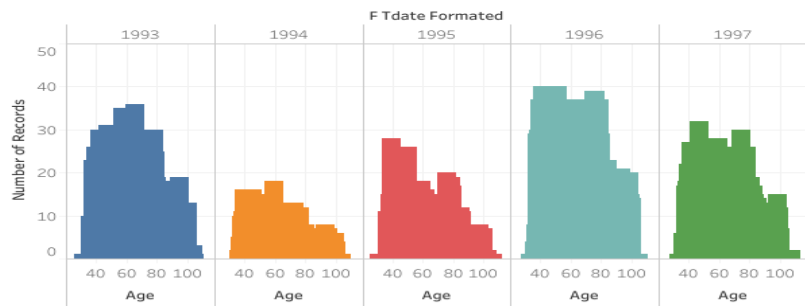As there is huge amount of transaction we can se most of the avg amount is around 2K .

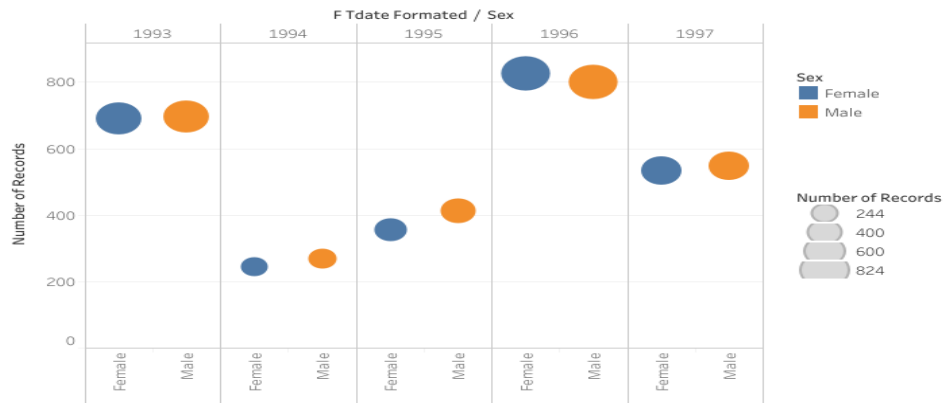# Interactive Dashboard - Tableau

Logical Steps and code snippet

- **Create a dashboard including all Worksheet in Tableau**

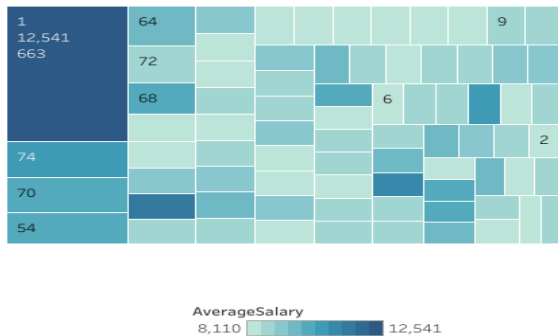**Tableau Online link :https://eu-west-1a.online.tableau.com/#/site/sudipbairagi/views/BankCaseStudy/Dashboard1?:iid=1**