Comprehensive Report

on

"Data Analytics and Data Visualization with Tableau"

(Summer Course)

Submitted by:

Name: **Rahul S Bhat**

SRN: **PES2UG19CS315**

SEM: **7th**

Faculty Handling the Course:

Dr.Sudeepa Roy Dey
Dr.Prajwala TR
Prof.Ruby Dinakar

4th -9th, July 2022

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
PES UNIVERSITY, EC campus
(Established under Karnataka Act No. 16 of 2013)
Electronic City, Hosur Road, Bengaluru – 560 100, Karnataka, Ind

Table of Contents

Title	Page No.
1. About the Dataset	2
2. Preprocessing	3
3. Weka Statistics	6
4. Measures and Dimensions	7
5. Context Filter	8
6. Visualization in Tableau	9
7. The Final Dashboard	12

About The Dataset:

Dataset: Indian Card Payment Dataset

Source: Kaggle

No. of columns: 20

No. of Rows: 5592

The dataset contains bank wise payment statistics on monthly basis released by RBI. It contains statistics like the no of debit cards, no. of credit cards, no. of ATMs, no. of pos, Bank names, year, month, taxation details, etc. and the data is spread across 8 years i.e. from Apr'2011 to Aug'2019. This data can be used to predict the payment trend in India.

Preprocessing:

```
In [1]: import pandas
```

```
In [2]: df=pandas.read_csv('C:/Users/Rahul Bhat/Downloads/archive (7)/rbi_payment_data_as_on_aug_2019.csv')
```

```
In [3]: df.head()
```

Out[3]:

	month	year	month_number	start_date	end_date	bank_name	no_atms_on_site	no_atms_off_site	no_pos_on_line	no_pos_off_line	no_credit_cards	no_c
0	November	2011	11	2011-11-01	2011-11-30	Allahabad Bank	207	109	0	0.0	0.0	
1	November	2011	11	2011-11-01	2011-11-30	Andhra Bank	479	554	2122	0.0	121514.0	
2	November	2011	11	2011-11-01	2011-11-30	Bank of Baroda	1242	580	4332	0.0	70776.0	
3	November	2011	11	2011-11-01	2011-11-30	Bank of India	838	792	1930	501.0	119248.0	
4	November	2011	11	2011-11-01	2011-11-30	Bank of Maharashtra	359	141	77	404.0	23436.0	

The dataset contained around 183 null values which were dropped because it constituted a small fraction of dataset.

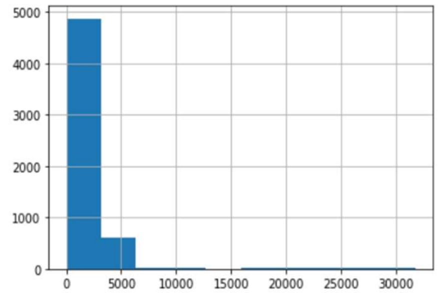
```
In [5]: df.dropna(inplace=True)
```

```
In [6]: df.isnull().sum()
```

```
Out[6]: month                                 0
        year                                  0
        month_number                          0
        start_date                            0
        end_date                              0
        bank_name                             0
        no_atms_on_site                       0
        no_atms_off_site                      0
        no_pos_on_line                        0
        no_pos_off_line                       0
        no_credit_cards                       0
        no_credit_card_atm_txn                0
        no_credit_card_pos_txn                0
        no_credit_card_atm_txn_value_in_mn    0
        no_credit_card_pos_txn_value_in_mn    0
        no_debit_cards                        0
        no_debit_card_atm_txn                 0
        no_debit_card_pos_txn                 0
        no_debit_card_atm_txn_value_in_mn     0
        no_debit_card_pos_txn_value_in_mn     0
        dtype: int64
```

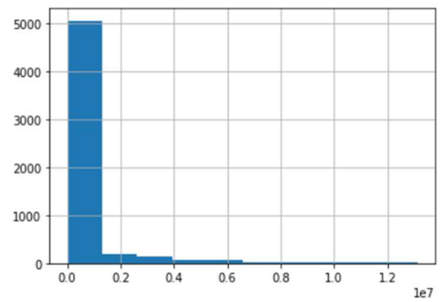
In [7]:
```
df['no_atms_on_site'].hist()
```

Out[7]: <AxesSubplot:>



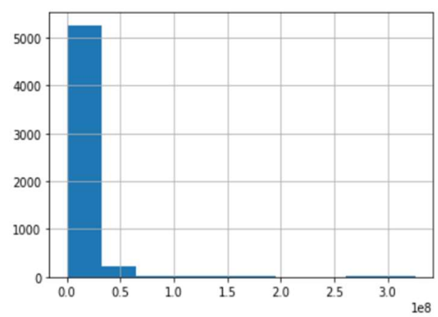
In [8]:
```
df['no_credit_cards'].hist()
```

Out[8]: <AxesSubplot:>



In [9]:
```
df['no_debit_cards'].hist()
```

Out[9]: <AxesSubplot:>



In [10]:
```
for i in df:
    print(df[i].is_monotonic)
```

```
False
False
False
False
False
False
False
False
False
False
False
False
False
False
False
False
False
False
False
False
```

In [11]:
```
df.corr(method='pearson')
```
Out[11]:

	year	month_number	no_atms_on_site	no_atms_off_site	no_pos_on_line	no_pos_off_line	no_credit_cards	no_credit_cai
year	1.000000	-0.118098	0.107228	0.084806	0.184302	-0.133035	0.102894	
month_number	-0.118098	1.000000	-0.000035	-0.000602	-0.004781	-0.008199	-0.002087	
no_atms_on_site	0.107228	-0.000035	1.000000	0.894283	0.555805	0.083749	0.503003	
no_atms_off_site	0.084806	-0.000602	0.894283	1.000000	0.720520	0.062347	0.634588	
no_pos_on_line	0.184302	-0.004781	0.555805	0.720520	1.000000	0.106744	0.789484	
no_pos_off_line	-0.133035	-0.008199	0.083749	0.062347	0.106744	1.000000	0.105951	
no_credit_cards	0.102894	-0.002087	0.503003	0.634588	0.789484	0.105951	1.000000	
no_credit_card_atm_txn	0.153989	0.001785	0.555357	0.654139	0.790412	-0.013969	0.907061	
no_credit_card_pos_txn	0.157263	0.001883	0.442140	0.558907	0.733748	0.048192	0.955653	
no_credit_card_atm_txn_value_in_mn	-0.040108	-0.023046	0.068970	-0.011715	-0.008149	0.171548	-0.007131	
no_credit_card_pos_txn_value_in_mn	-0.001668	0.016092	0.123356	0.042012	0.072357	0.187836	0.089538	
no_debit_cards	0.141143	0.001056	0.927416	0.906335	0.556781	-0.000187	0.464452	
no_debit_card_atm_txn	0.057558	0.000551	0.916650	0.920372	0.546567	0.011324	0.483949	
no_debit_card_pos_txn	0.218526	0.003995	0.778889	0.814315	0.742469	0.060802	0.718238	
no_debit_card_atm_txn_value_in_mn	-0.034470	0.001264	0.212515	0.026167	0.020376	0.325116	0.017026	
no_debit_card_pos_txn_value_in_mn	-0.040999	-0.021489	0.122726	-0.007246	-0.002974	0.241322	-0.004166	

In [12]:
```
df.describe()
```
Out[12]:

	year	month_number	no_atms_on_site	no_atms_off_site	no_pos_on_line	no_pos_off_line	no_credit_cards	no_credit_card_atm_txn	no_credit_card_p
count	5588.000000	5588.000000	5588.000000	5588.000000	5588.000000	5588.000000	5.588000e+03	5588.000000	5.5880
mean	2015.075340	6.407659	1545.015927	1411.407122	30503.386364	115.342341	4.684350e+05	8163.134216	1.3716
std	2.491269	3.364588	3068.939368	3665.141817	92442.167634	829.432634	1.391089e+06	23789.579734	4.5636
min	2011.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000e+00	0.000000	0.0000
25%	2013.000000	4.000000	168.000000	109.000000	0.000000	0.000000	0.000000e+00	0.000000	0.0000
50%	2015.000000	6.000000	692.000000	401.500000	1769.000000	0.000000	5.492500e+03	18.000000	7.9385
75%	2017.000000	9.000000	1686.000000	827.000000	10908.000000	0.000000	1.339610e+05	3946.250000	1.7509
max	2019.000000	12.000000	31749.000000	33209.000000	988458.000000	13945.000000	1.313833e+07	202780.000000	4.8011

In [16]:
```
len(df['bank_name'].unique())
```
Out[16]: 154

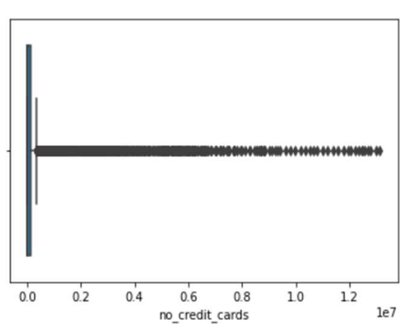
It was seen that the data wasn’t monotonic and Pearson correlation was applied

The dataset consisted data for 154 different banks across India.

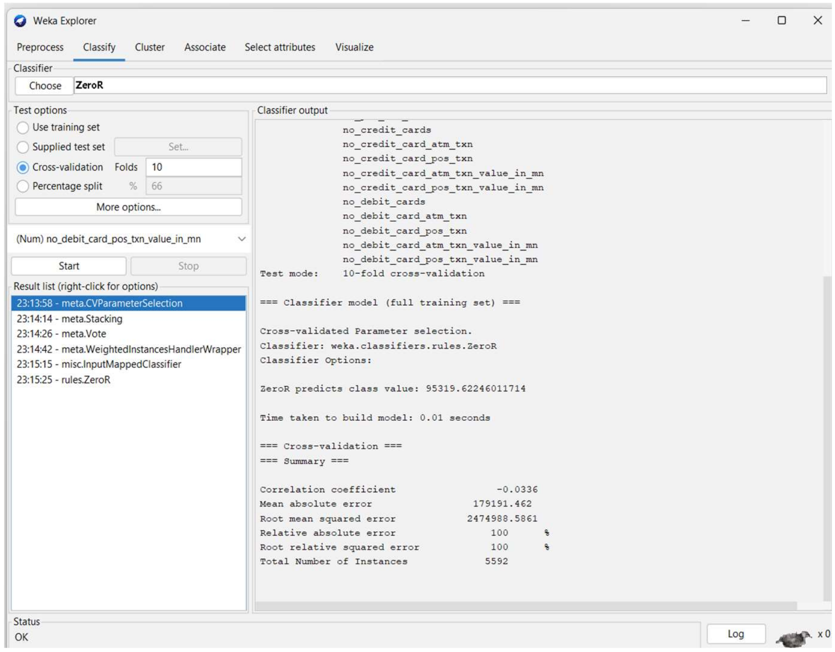
Histograms show a skewed distribution and hence the data is not normal.

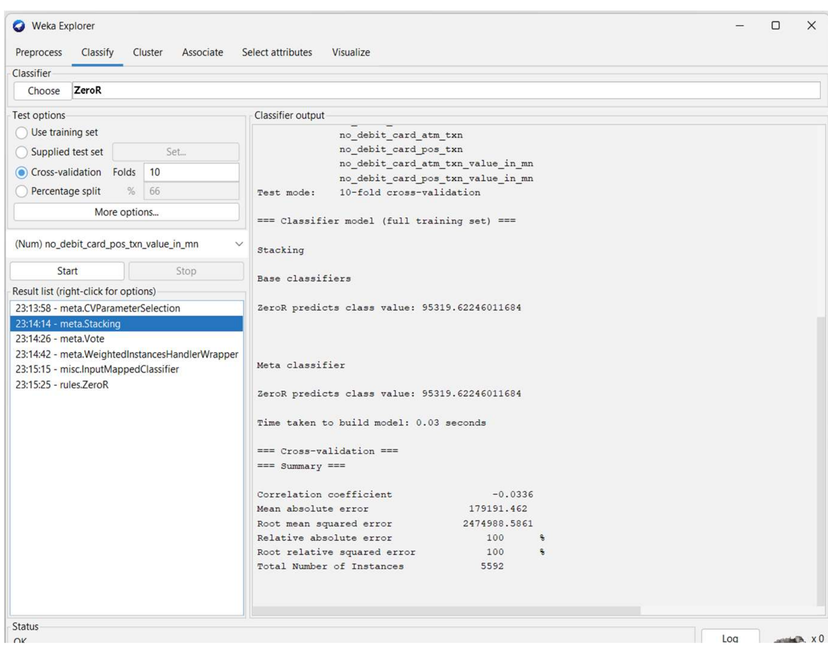
Data was too random

In [21]:
```
sns.boxplot(df['no_credit_cards'])
```
```
C:\Users\Rahul Bhat\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.9_qbz5n2kfra8p0\LocalCache\local-packages\Python3
9\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, t
he only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an err
or or misinterpretation.
  warnings.warn(
```
Out[21]: <AxesSubplot:xlabel='no_credit_cards'>



Weka Statistics





Measures and Dimensions

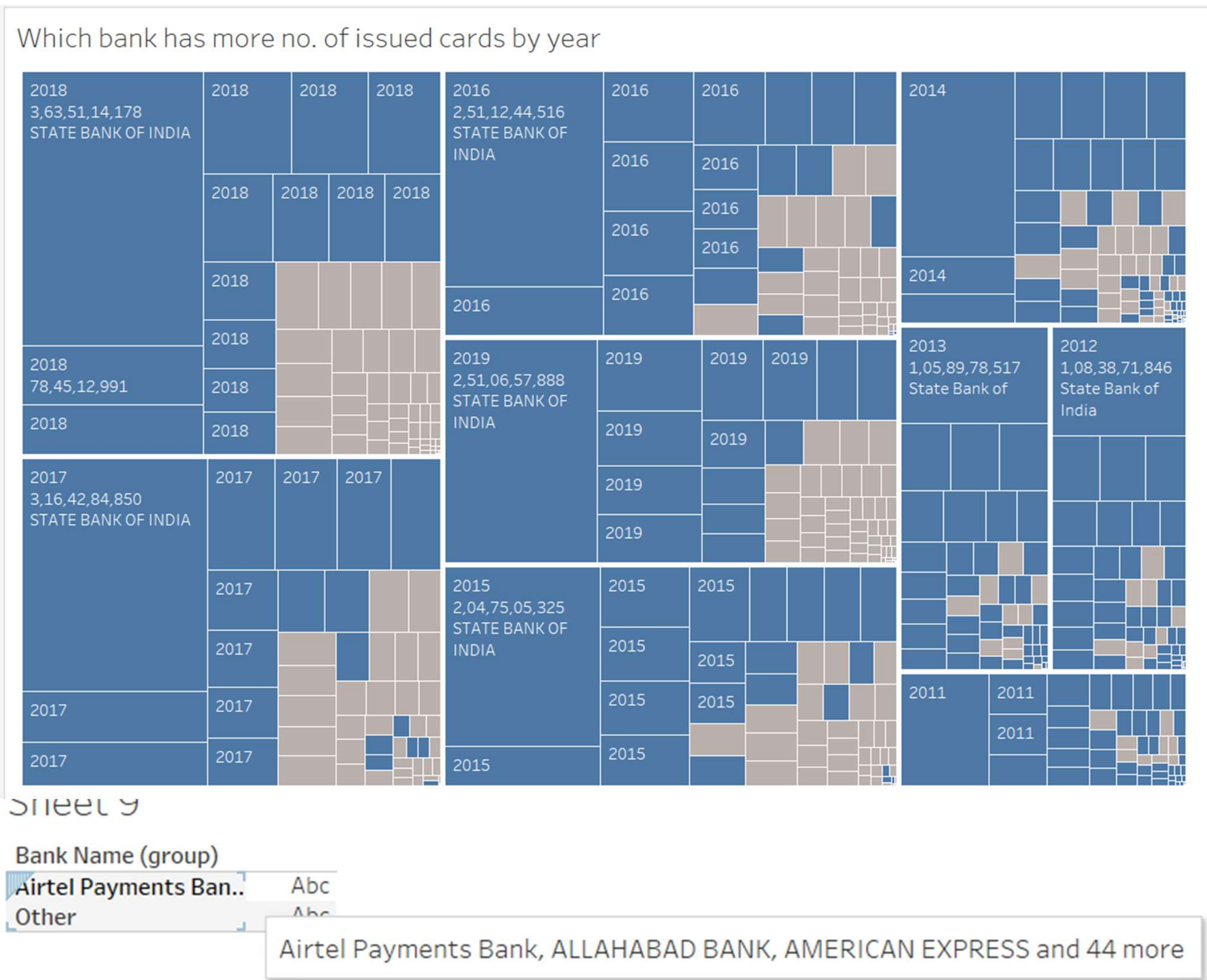
Dimensions:	Measures:
Bank name	No ATMs on site
Year	No ATMs off site
Month number	No POSs on site
Start date	No POSs off site
End date	No credit cards
	No debit cards
	rbi_payment_data_as_on_aug_2019(count) (generated measure)

Calculated fields/ Measures:

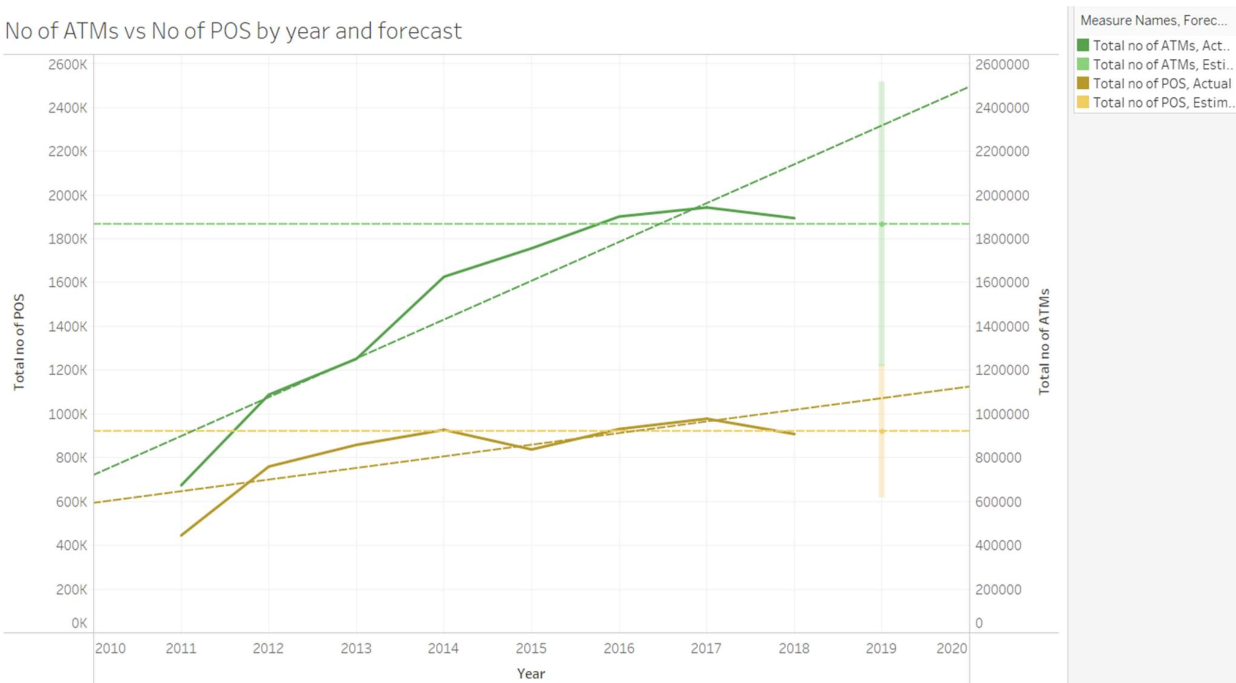
Total no. of ATMs = No. of Onsite + No. of offsite
Total no. of POS= No. of Onsite + No. of off site
Total no. of Cards= No. of credit cards + No. of debit cards

Context filters:

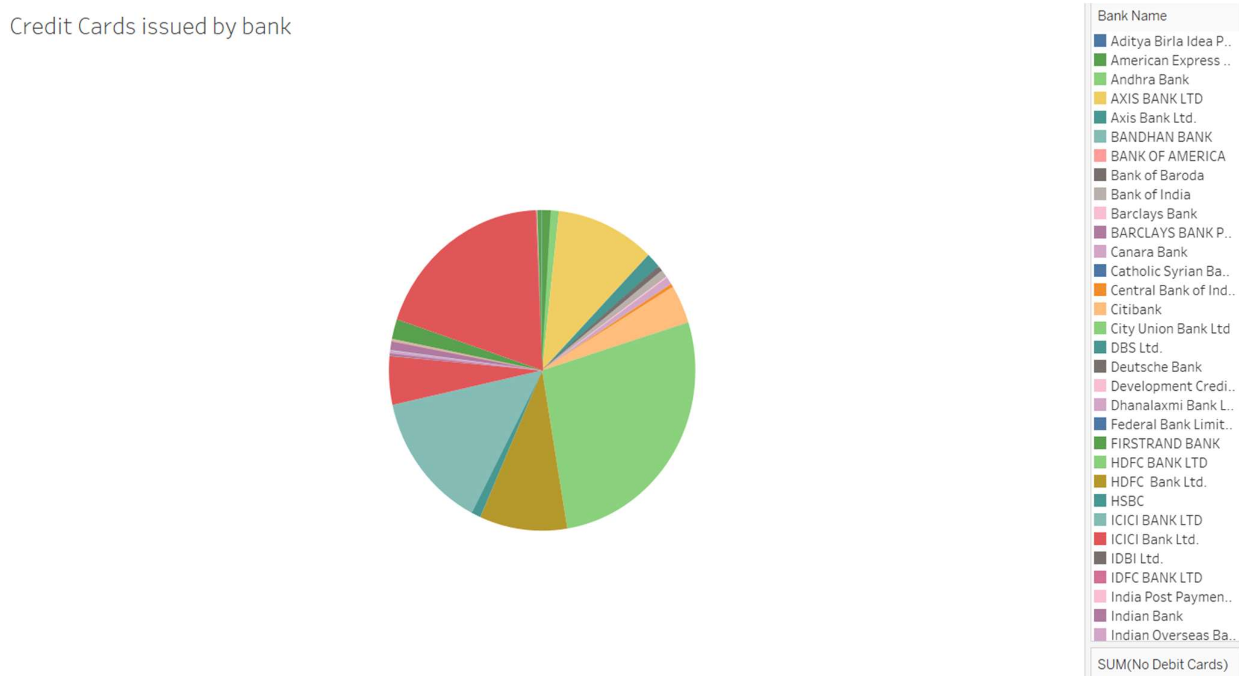
The less contributing banks to the no of cards were **grouped** and were excluded from the analysis using the context filter.



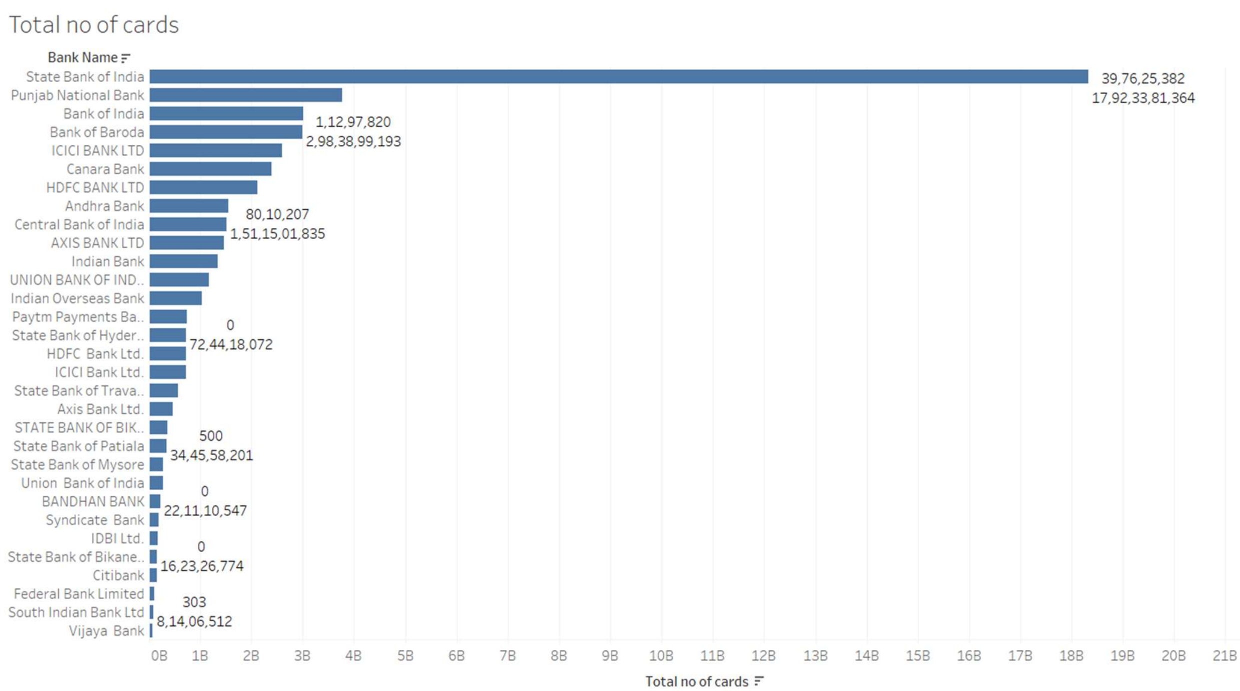
Visualization in Tableau



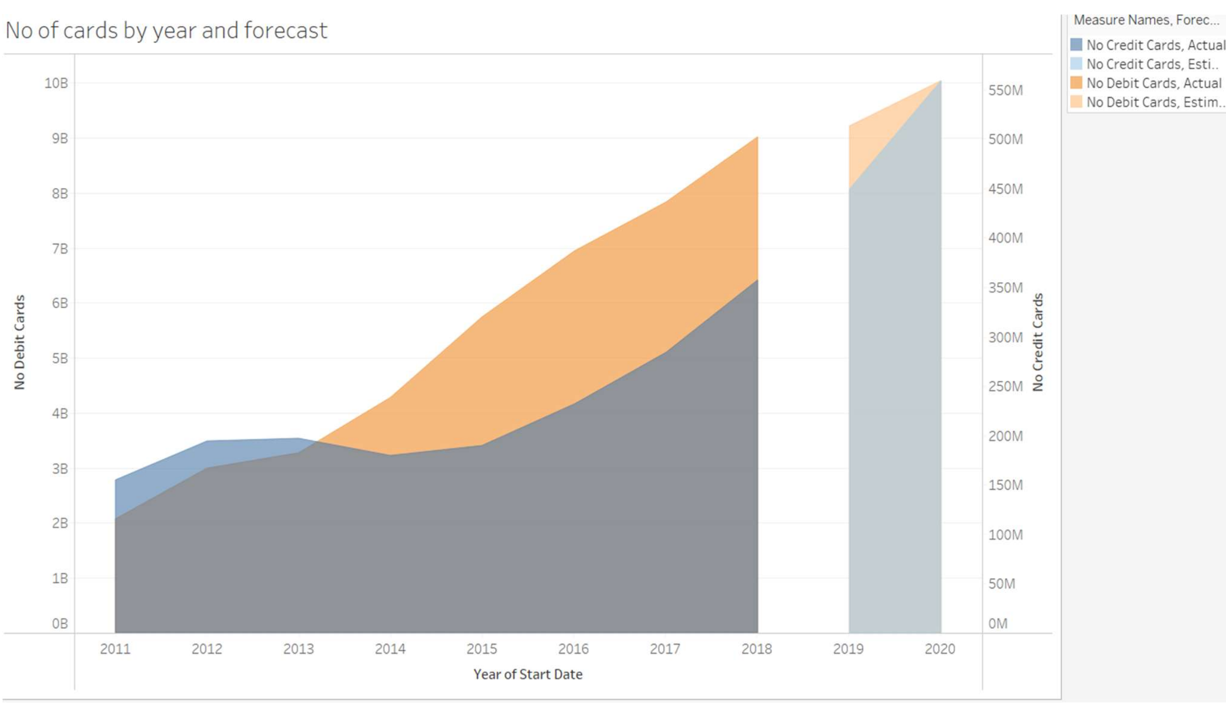
The no. of ATMs shows an increasing trend but the forecast shows a bit of depression and this might continue as a result of digitalization of payments.



Most number of credit cards were issued by HDFC bank and 2nd stands the State Bank of India

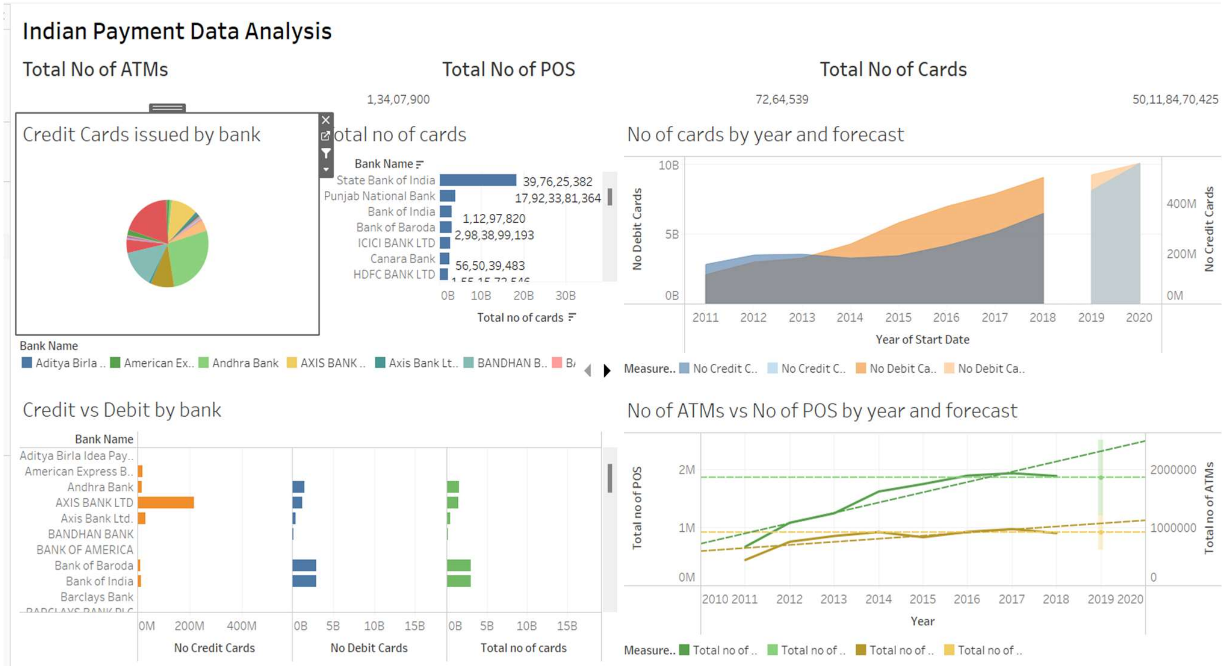


When you consider the total no. of cards, first stands the State Bank of India and the HDFC ranks seventh. Hence we can say that the despite lesser no. of credit cards, State Bank of India has issued more no. of debit cards than the HDFC bank.



The no. of both credit and debit cards issued is showing upward trend and a positive forecast but at some point of time it will saturate.

The Final Dashboard



Interaction has been enabled and the actions have been added to the pie chart. Pie chart is the control. By clicking on the particular sector in the pie chart, the visuals for that particular sector is generated.

The three stats on the top namely Total no of ATMs, Total no of POS, Total no of cards are also acted upon by this action.

Some more Insights from the dashboard:

The no of ATMs sees a linear increase trend but is forecasted as to saturate whereas the POS is almost saturated from the beginning. This might be the results of digitalization of payments that the stats seem to be saturating and will drop down in near future.

On comparing the stats about the cards issued it seems that HDFC is doing better compared to other banks in case of credit cards whereas State Bank of India is doing better when it comes to no of debit cards and total no of cards.

The dual plot of no of credit vs debit cards sees the same increasing trend is forecasted to increase further the next year.