PROJECT: CREDIT CARD CASE STUDY - SEGMENTATION

ABSTRACT:

The case requires developing a customer segmentation to define marketing strategy.
The sample dataset summarizes the usage behaviour of 9000 active credit card holder during the last 6 months.

SUMMARIES OF PROBLEM, DATA, METHODS, AND TECHNOLOGIES:

The project can be internally divided into four tasks namely –

A> Preparing data built with intelligent KPIs.
B> Provide the detailed insights/observations based on the analysis.
C> Cluster Analysis.
D> Providing strategic insights.

DATA SUMMARY

The input data provided is in CSV data format. The data need to be imported using ‘read_csv’ function of ‘pandas’ library.The data required data manipulation in terms of data conversion and cleaning.

METHODS SUMMARY

Table shows the wide variety of data pre-processing, analysis, and visualization techniques that I applied to complete the tasks as part of the project –

Task	Task Details	Analytical Techniques	Visualization Techniques
Data manipulation and preparation	1.Preparing data with derived intelligent KPIs. 2. Provide the detailed insights/observation sbased on the analysis	Descriptive statistics Straightforward ata manipulation	pandas_profiling.Profile Report matplotlib.pyplot (bar, scatter plot)
Modelling and Performance	Cluster Analysis. Providing strategic insights.	K-Means cluster analysis	

A> DATA PREPERATION AND DERIVING INTELLIGENT KPIs

a) Load the input file into a dataframe- credit and check for missing values.
b) Null values of variables credit_limit will be filled by median value and minnimum_payemnts with 0's

c) KPIs Derived

1. **Monthly Average Purchases** –

```
credit['MONTHLY_AVG_PURCHASE'] = credit['PURCHASES']/credit['TENURE']
```

2. **Monthly Cash Advance** –

```
credit['MONTHLY_AVG_CASH_ADVANCE'] = credit['CASH_ADVANCE']/credit['TENURE']
```

. 3. **Defining purchase type-**#There aree 4 types of purchase behavior - deriving categorical variables based on the behavior

```
def purchasetype(credit):
    if ((credit.ONEOFF_PURCHASES == 0) & (credit.INSTALLMENTS_PURCHASES == 0)):
        return 'NONE'
    if ((credit.ONEOFF_PURCHASES > 0) & (credit.INSTALLMENTS_PURCHASES == 0)):
        return 'ONE_OFF'
    if ((credit.ONEOFF_PURCHASES > 0) & (credit.INSTALLMENTS_PURCHASES > 0)):
        return 'ONEOFF_INSTALLMENT'
    if ((credit.ONEOFF_PURCHASES == 0) & (credit.INSTALLMENTS_PURCHASES > 0)):
        return 'INSTALLMENTS'
credit['PURCHASE_TYPE']=credit.apply(purchasetype,axis=1 )
```

4. **LIMIT USAGE** (Lower score implies customers are maintaining their balance properly)

```
credit['LIMIT_USAGE'] = credit.apply(lambda x: x['BALANCE']/x['CREDIT_LIMIT'],axis =1)
```

5. **PAYMENT_MINPAYMENT-** The where clause is being used to avoid div by zero error

```
credit['PAYMENT_MINPAYMENT'] = np.where(credit['MINIMUM_PAYMENTS']== 0,
credit['PAYMENTS'], credit['PAYMENTS']/credit['MINIMUM_PAYMENTS'])
```

d) Drop the variables used to generate KPI's.

e) Outliers are handled.

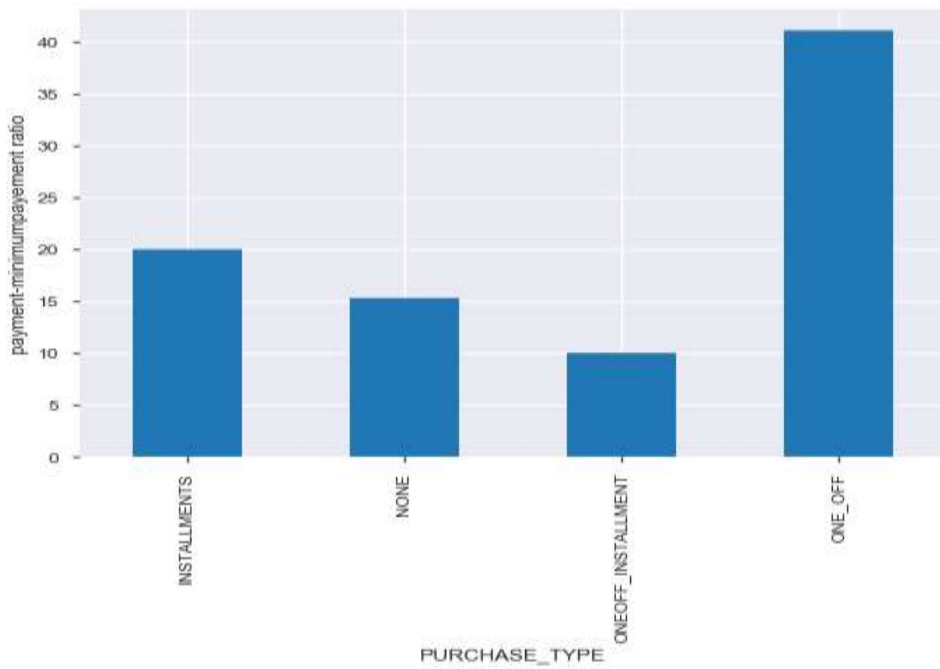
f.) Feature encoding for categorical variable PURCHASE_TYPE using on hot encoding.

g) Final dataset is *card_ new* dataframe

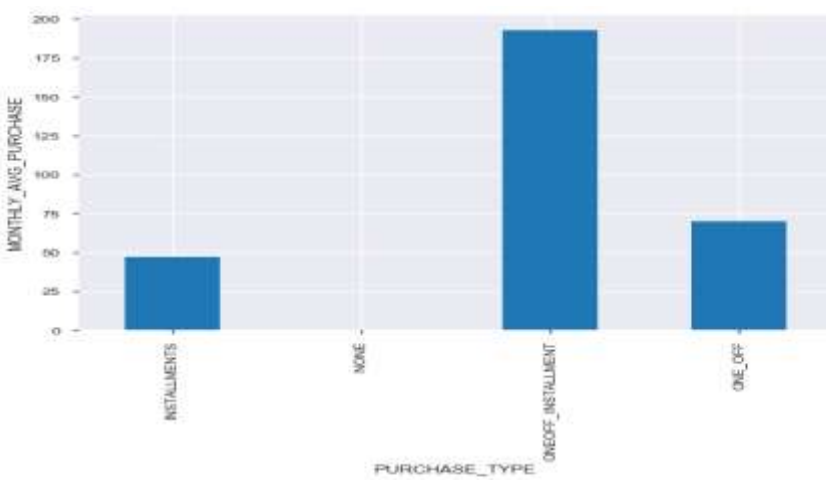
B>Provide the detailed insights/observations based on the analysis :

- Customers with one off and installments are paying their dues
- Customers with one off and installments do most monthly purchases
- Customers with no one off and installments take more monthly cash advance
- Customers with installments have good credit score

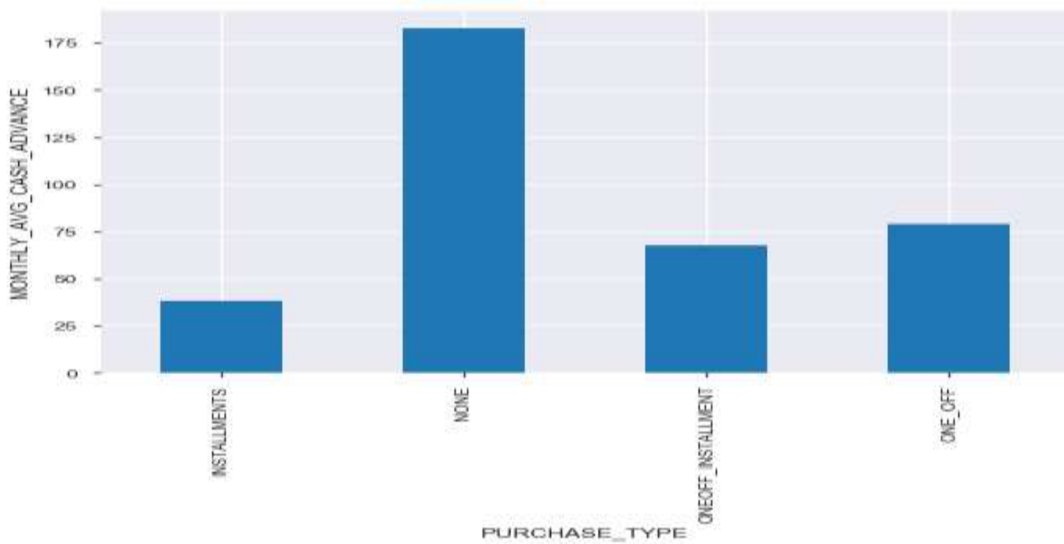
PLOT 1: PURCHASE_TYPE VS PAYMENT-MINIMUMPAYMENT RATIO



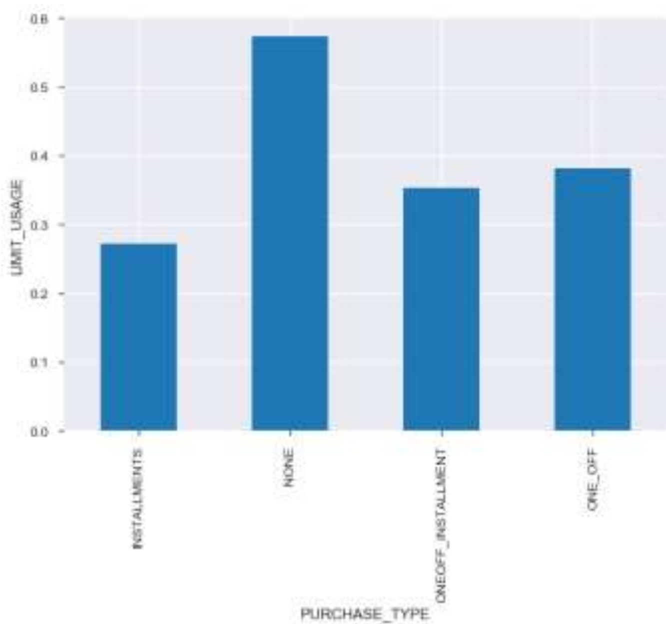
PLOT 2: PURCHASE_TYPE VS MONTHLY_AVG_PURC HASE



PLOT 3: PURCHASE_TYPE VS MONTHLY_CASH ADVANCE



PLOT 4: PURCHASE_TYPE VS LIMIT_USAGE



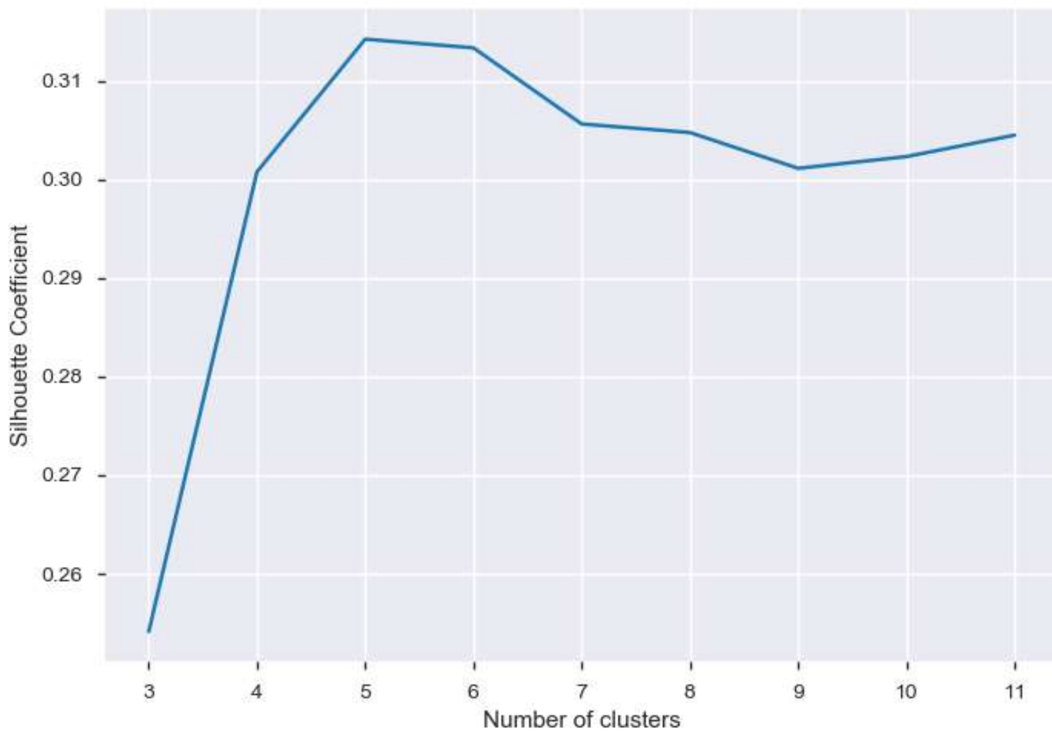
MODEL BUILDING AND PERFORMANCE

K MEANS MODEL IS USED FOR CLUSTERING

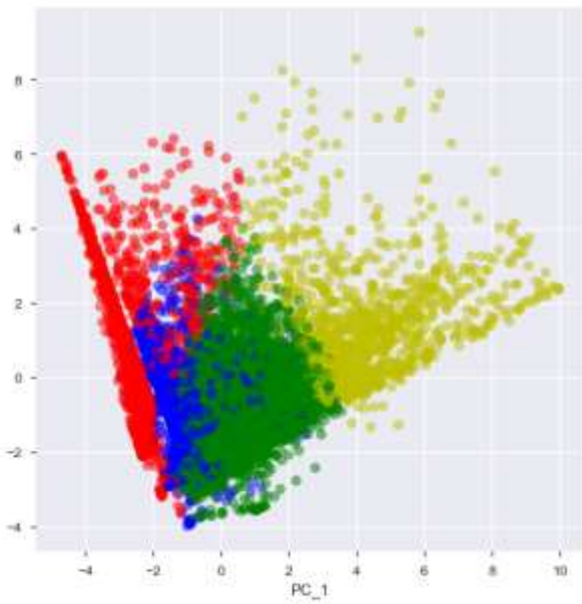
C> Cluster Analysis.

- Correlation among variable is checked to remove multicoliearity.

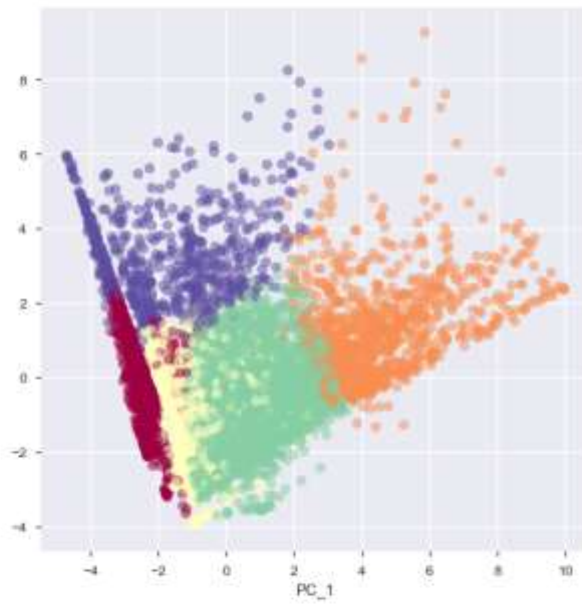
- Standardizing the data and applying PCA to get the optimal component. Since 7 components are explaining 85% of variation so we select 7 component.
    - *np.cumsum(np.round(pc.explained_variance_ratio_, decimals=4)*100)*
    - `array([32.73, 49.57, 60.19, 69.18, 75.54, 80.61, 85.02, 88.7 , 91.05, 93.21, 95.31, 97.26, 98.37, 99.05, 99.68, 99.88, 99.98])`
- Using Silhouette Coefficient to get optimal cluster - The solution can be 4 or 5 or 6



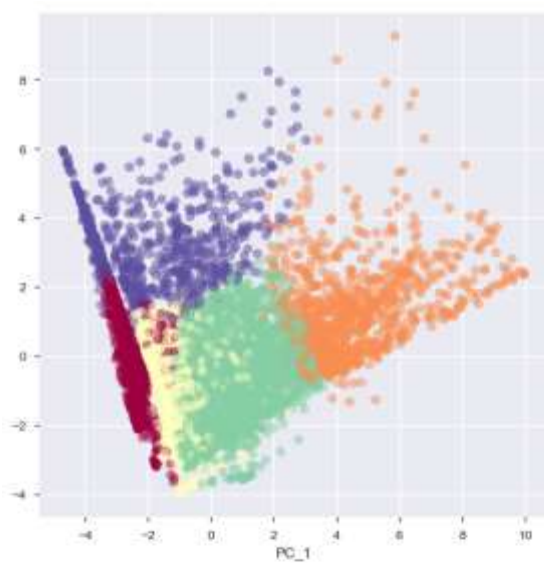
CLUSTER SOLUTIONS



CLUSTER 4



CLUSTER 5



CLUSTER 6

D> Providing strategic insights.

Overall profiling for K Means cluster solution

		Overall	KM4_1	KM4_2	KM4_3	KM4_4	KM5_1	KM5_2	KM5_3	KM5_4	KM5_5	KM6_1	KM6_2	KM6_3	KM6_4	KM6_5	KM6_6
Seg_Pct		100%	27%	19%	41%	13%	21%	11%	19%	40%	9%	20%	11%	8%	10%	16%	35%
PURCHASES_TRX	AVG	14.15	1.20	6.22	13.98	54.38	0.12	56.37	5.90	14.66	8.76	0.17	4.22	9.04	58.68	6.70	16.61
MONTHLY_AVG_PURCHASE	AVG	80.56	6.87	57.37	60.26	338.13	0.40	357.11	53.64	63.45	50.35	0.63	26.01	51.50	377.08	59.15	72.12
CASH_ADVANCE_TRX	AVG	3.08	7.40	1.89	0.91	2.53	4.22	1.59	1.35	0.77	16.63	4.45	0.51	17.06	1.59	1.71	0.93
MONTHLY_AVG_CASH_ADVANCE	AVG	84.92	208.33	53.00	24.08	63.22	129.44	41.03	38.38	20.61	431.58	132.85	22.97	443.75	42.15	47.03	23.93
LIMIT_USAGE	AVG	0.39	0.58	0.37	0.28	0.33	0.57	0.31	0.34	0.28	0.62	0.62	0.02	0.61	0.31	0.41	0.33
PAYMENT_MINPAYMENT	AVG	6.67	5.14	6.42	6.39	11.20	5.46	11.82	7.02	6.23	4.04	2.67	24.18	4.13	12.49	3.73	3.87
PURCHASE_TYPE_ONEOFF_INSTALLMEN	%YES	0.31	0.06	0.00	0.42	0.95	0.00	0.94	0.00	0.44	0.31	0.00	0.09	0.32	0.94	0.00	0.50
PURCHASE_TYPE_NONE	%YES	0.23	0.84	0.00	0.00	0.00	0.95	0.00	0.00	0.00	0.32	0.94	0.16	0.33	0.00	0.00	0.00
PURCHASE_TYPE_ONE_OFF	%YES	0.21	0.04	1.00	0.00	0.04	0.00	0.05	0.96	0.00	0.24	0.00	0.24	0.23	0.05	1.00	0.00
CREDIT_LIMIT	AVG	4474.25	4391.41	4265.63	3712.16	7401.36	3608.33	7462.88	4154.30	3722.22	6805.91	3642.62	3746.83	6953.71	7661.82	4131.06	3819.98
20% or less than overall avg	20% or less than overall avg																

Based on segment distribution, profiling and SC scoring best solution is for 5 category segmentation

1. cluster 1 -- It is 21 % of total customer base and customer have low monthly purchase, but high monthly cash advance and poor credit score, This group is inclined to No purchases of any type
2. cluster 2-- It is 11 % of total customer base and customer have high monthly purchase, low monthly cash advance and good credit score.This group is doing BOTH one off and installment purchases
3. cluster 3: -- 19 % of customer base. this group has monthly purchase between customer 1 and 2 and monthly cash advace and crdit score in comparision to cluster 2. customers of this group are doing mostly one_off purchases.
4. cluster 4-It is 40 % of total customer base . This group has best limit usage (lowest) and doing moderate one off and installment purchase.
5. cluster 5 --It is 9 % of total customer base . this group has lower monthly purchase than cluster 1 but higher cash advance than cluster 1. customer are inclined to No or one off purchases