

# Assignment 3 - Santhosh

March 18, 2024

```
[1]: import pandas as pd
import numpy as np
pd.plotting.register_matplotlib_converters()
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import plotly.express as px
from datetime import datetime
```

```
[4]: creditcard='Credit card transactions - India - Simple.csv'
df = pd.read_csv(creditcard,index_col="index")
df.head()
```

```
[4]:
```

|       | City                  | Date      | Card Type | Exp Type | Gender | Amount |
|-------|-----------------------|-----------|-----------|----------|--------|--------|
| index |                       |           |           |          |        |        |
| 0     | Delhi, India          | 29-Oct-14 | Gold      | Bills    | F      | 82475  |
| 1     | Greater Mumbai, India | 22-Aug-14 | Platinum  | Bills    | F      | 32555  |
| 2     | Bengaluru, India      | 27-Aug-14 | Silver    | Bills    | F      | 101738 |
| 3     | Greater Mumbai, India | 12-Apr-14 | Signature | Bills    | F      | 123424 |
| 4     | Bengaluru, India      | 5-May-15  | Gold      | Bills    | F      | 171574 |

```
[5]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 26052 entries, 0 to 26051
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  -
0   City        26052 non-null  object
1   Date        26052 non-null  object
2   Card Type   26052 non-null  object
3   Exp Type    26052 non-null  object
4   Gender      26052 non-null  object
5   Amount      26052 non-null  int64
dtypes: int64(1), object(5)
memory usage: 1.4+ MB
```

```
[6]: from datetime import datetime
df['Date'] = pd.to_datetime(df['Date'])
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 26052 entries, 0 to 26051
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  -
0   City         26052 non-null  object
1   Date         26052 non-null  datetime64[ns]
2   Card Type    26052 non-null  object
3   Exp Type     26052 non-null  object
4   Gender       26052 non-null  object
5   Amount       26052 non-null  int64
dtypes: datetime64[ns](1), int64(1), object(4)
memory usage: 1.4+ MB
```

```
[7]: df.nunique()
```

```
[7]: City          986
Date            600
Card Type        4
Exp Type         6
Gender           2
Amount         24972
dtype: int64
```

```
[8]: df['date'] = df['Date'].dt.strftime('%d')
df['Year'] = df['Date'].dt.strftime('%Y')
df['Month'] = df['Date'].dt.strftime('%B')
df['Month_Year'] = df['Date'].dt.strftime('%B-%Y')
df
```

```
[8]:
```

|       | City                  | Date       | Card Type | Exp Type | Gender | Amount | \ |
|-------|-----------------------|------------|-----------|----------|--------|--------|---|
| index |                       |            |           |          |        |        |   |
| 0     | Delhi, India          | 2014-10-29 | Gold      | Bills    | F      | 82475  |   |
| 1     | Greater Mumbai, India | 2014-08-22 | Platinum  | Bills    | F      | 32555  |   |
| 2     | Bengaluru, India      | 2014-08-27 | Silver    | Bills    | F      | 101738 |   |
| 3     | Greater Mumbai, India | 2014-04-12 | Signature | Bills    | F      | 123424 |   |
| 4     | Bengaluru, India      | 2015-05-05 | Gold      | Bills    | F      | 171574 |   |
| ...   | ...                   | ...        | ...       | ...      | ...    | ...    |   |
| 26047 | Kolkata, India        | 2014-06-22 | Silver    | Travel   | F      | 128191 |   |
| 26048 | Pune, India           | 2014-08-03 | Signature | Travel   | M      | 246316 |   |
| 26049 | Hyderabad, India      | 2015-01-16 | Silver    | Travel   | M      | 265019 |   |
| 26050 | Kanpur, India         | 2014-09-14 | Silver    | Travel   | M      | 88174  |   |
| 26051 | Hyderabad, India      | 2013-10-19 | Signature | Bills    | M      | 184410 |   |

```

      date  Year      Month      Month_Year
index
0      29  2014    October    October-2014
1      22  2014     August    August-2014
2      27  2014     August    August-2014
3      12  2014     April     April-2014
4      05  2015      May      May-2015
...    ...    ...      ...      ...
26047  22  2014      June     June-2014
26048  03  2014     August    August-2014
26049  16  2015    January    January-2015
26050  14  2014    September  September-2014
26051  19  2013    October    October-2013

```

[26052 rows x 10 columns]

```
[9]: df[['City Name', 'Country']] = df.City.str.split(",", expand = True)
df
```

```
[9]:
      City      Date  Card Type  Exp Type  Gender  Amount \
index
0      Delhi, India 2014-10-29    Gold   Bills      F    82475
1  Greater Mumbai, India 2014-08-22  Platinum  Bills      F    32555
2      Bengaluru, India 2014-08-27    Silver   Bills      F   101738
3  Greater Mumbai, India 2014-04-12  Signature  Bills      F   123424
4      Bengaluru, India 2015-05-05     Gold   Bills      F   171574
...    ...      ...      ...      ...      ...
26047      Kolkata, India 2014-06-22    Silver  Travel      F   128191
26048      Pune, India 2014-08-03  Signature  Travel      M   246316
26049  Hyderabad, India 2015-01-16    Silver  Travel      M   265019
26050      Kanpur, India 2014-09-14    Silver  Travel      M    88174
26051  Hyderabad, India 2013-10-19  Signature  Bills      M   184410

```

```

      date  Year      Month      Month_Year      City Name  Country
index
0      29  2014    October    October-2014      Delhi    India
1      22  2014     August    August-2014  Greater Mumbai  India
2      27  2014     August    August-2014      Bengaluru  India
3      12  2014     April     April-2014  Greater Mumbai  India
4      05  2015      May      May-2015      Bengaluru  India
...    ...    ...      ...      ...      ...
26047  22  2014      June     June-2014      Kolkata    India
26048  03  2014     August    August-2014      Pune      India
26049  16  2015    January    January-2015  Hyderabad    India
26050  14  2014    September  September-2014  Kanpur      India
26051  19  2013    October    October-2013  Hyderabad    India

```

[26052 rows x 12 columns]

```
[10]: df.describe().T
```

```
[10]:
```

|        | count    | mean          | std           | min    | 25%      | 50% \    |
|--------|----------|---------------|---------------|--------|----------|----------|
| Amount | 26052.0  | 156411.537425 | 103063.254287 | 1005.0 | 77120.25 | 153106.5 |
|        |          | 75%           | max           |        |          |          |
| Amount | 228050.0 |               | 998077.0      |        |          |          |

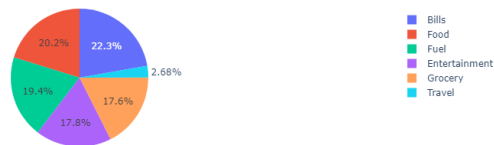
```
[11]: px.pie(df.sort_values(by='Amount',ascending=1), values='Amount', names='Card_
↳Type', title='Proportion Card Types',
color_discrete_sequence=["grey", "green", "lightgrey", "gold"]).show()
```

Proportion Card Types



```
[12]: px.pie(df.sort_values(by='Amount',ascending=1), values='Amount', names='Exp_
↳Type', title='Proportion of Exp Types').show()
```

Proportion of Exp Types



```
[ ]:
```