***Report of Personal Loan and Insight***

**Introduction**

The data is all about the personal loan. In loan process there are different stages which was indicated by stage column in the data. City wise Income is also given in the data.

My process to find the result and Analysis the data

1. Importing all the necessary library according to the data.

* import numpy as np
* import pandas as pd
* import matplotlib.pyplot as plt
* import seaborn as sns

data = pd.read\_csv('https://raw.githubusercontent.com/akjadon/Finalprojects\_DS/master/Personal-Loan-Stats-and-Insights/assignment.csv')

data.head()

| **Id** | **Application Date** | **City** | **Income** | **Stage** |
| --- | --- | --- | --- | --- |
| **0** | 1 | 30-Sep-18 | North West Delhi | 307296 | Stage 1 |
| **1** | 2 | 30-Sep-18 | North West Delhi | 323976 | Stage 0 |
| **2** | 3 | 30-Sep-18 | Mumbai | 235632 | Stage 1 |
| **3** | 4 | 30-Sep-18 | Jaipur | 250668 | Stage 0 |
| **4** | 5 | 30-Sep-18 | Lakhimpur | 533616 | Stage 0 |

* data.columns.tolist()

['Id', 'Application Date', 'City', 'Income', 'Stage']

* Finding the Max value in the dataset

data.max

Id 10678

Application Date 31-Aug-18

City Yavatmal

Income 431716200

Stage Stage 5

dtype: object

* Data.min()

Id 1 Application

Date 01-Aug-18

City Adilabad

Income 204

Stage Stage 0

dtype: object

* data\_city=data.City
* mylist = list(dict.fromkeys(data\_city))
* mylist

'North West Delhi',

'Mumbai', 'Jaipur', 'Lakhimpur', 'Pune', 'Bengaluru', 'Gandhi Nagar', 'Nagpur', 'Chennai', 'Hyderabad', 'Mysuru', 'Ahmedabad', 'Other', 'Korba', 'Betul', 'Surat', 'Jabalpur', 'K.V.Rangareddy', 'Vijayapura', 'Gopalganj',

* len(mylist)

417

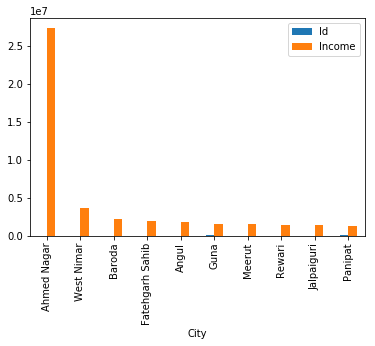
* data\_city.count()

10678

* Finding the top 10 City with higher Income.

city\_data = data.groupby('City').mean()

city\_data.sort\_values('Income',ascending=False).head(10).plot(kind = 'bar')



* print('The number of cities in entire dataset: ', len(data.City.unique()))
* print('The number of cities in only fifth Stage : ', len(data[data.Stage == 'Stage 5'].City.unique()))
* print('The number of cities in only Zero Stage : ', len(data[data.Stage == 'Stage 0'].City.unique()))
* print('The number of cities in only First Stage : ', len(data[data.Stage == 'Stage 1'].City.unique()))
* print('The number of cities in only Second Stage : ', len(data[data.Stage == 'Stage 2'].City.unique()))
* print('The number of cities in only Third Stage : ', len(data[data.Stage == 'Stage 3'].City.unique()))
* print('The number of cities in only Fourth Stage : ', len(data[data.Stage == 'Stage 4'].City.unique()))

The number of cities in entire dataset: 417

The number of cities in only fifth Stage : 4

The number of cities in only Zero Stage : 353

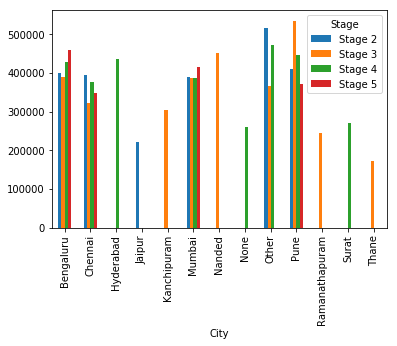
The number of cities in only First Stage : 333

The number of cities in only Second Stage : 6

The number of cities in only Third Stage : 9

The number of cities in only Fourth Stage : 8

* pd.pivot\_table(data[(data.Stage != 'Stage 0') & (data.Stage != 'Stage 1')], values= 'Income', index = 'City', columns= 'Stage', aggfunc= 'mean', fill\_value= 0).plot(kind = 'bar')



This Graph shows the main analysis of the given data as only 4 Cities have all the stages i. e Bengaluru, Chennai, Mumbai and Pune are having all the five stages means they are having the customer who waits for the personal loan given by bank.

* data.Income.describe()

count 1.067800e+04

mean 4.151758e+05

std 4.209443e+06

min 2.040000e+02

25% 1.962000e+05

50% 2.607120e+05

75% 3.765870e+05

max 4.317162e+08

Name: Income, dtype: float64

* data.Income.mean()

415175.8089529874

* data.median()

Id 5339.5

Income 260712.0

dtype: float64

* Creating the list of top 5 City based on the number of times occurred.

| **City** | **top5** |
| --- | --- |
| **0** | Bengaluru | 3594 |
| **1** | Mumbai | 1113 |
| **2** | Chennai | 1109 |
| **3** | Pune | 911 |
| **4** | Thane | 301 |

* Bengaluru is the city in which Personal loans are taken most of the time then after **Mumbai.**