

# Data Analysis Report

Loan Predictor

by Zhaklin Yanakieva

## **Overview:**

The data analysis report is to provide the findings of the research that was conducted during the collection of the data for this project.

## **Background:**

In the DataCollectionLedger file is explained more about how and from where the data was collected. In the proposal of the project is explained the goal and for what models the data will be used.

## **Purpose of the extraction effort:**

Or a needs assessment might be done to confirm the existence of needs or clarify the nature of needs that others have perceived. A job or task analysis might be undertaken to collect information directly related to the nature of the interactive multimedia product under development.

## **Data integration:**

In the project, I decided to combine data from two different csv files in order to gather more information. After that, data integration was applied in order to gather the data from two different csv files. The method that was used was 'Union', which makes me use exactly the same data characteristics for every column (type, domain and cast) so as to combine the datasets.

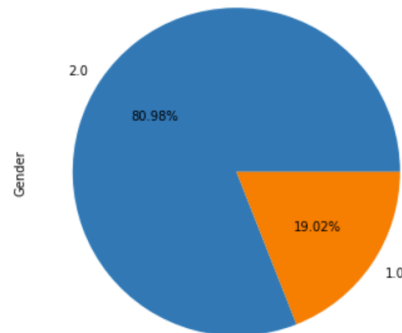
## **Methods:**

There are many ways to retrieve data from a source you selected. Common ways of extracting are from a file (csv or txt), from a database, JSON, API or web-scraping. In my project, I chose to extract

the data from two different csv files. I first check each file for the needed information about the loans and then, save the data together.

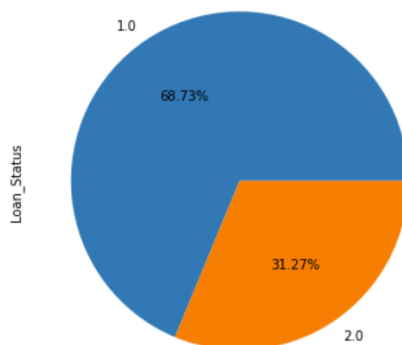
## Results:

```
[ ] df['Gender'].value_counts().plot(kind='pie', autopct='%1.2f%%', figsize=(6, 6))  
<matplotlib.axes._subplots.AxesSubplot at 0x7f99e0c59f50>
```

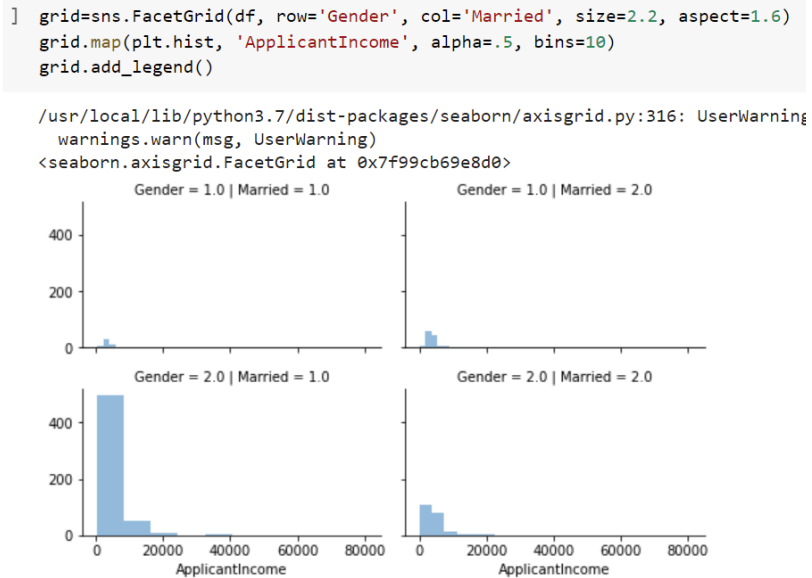


The pie chart above represents the percentage of the gender value. It is noticeable that the percentage of males who applied for a loan is greater than the one of females.

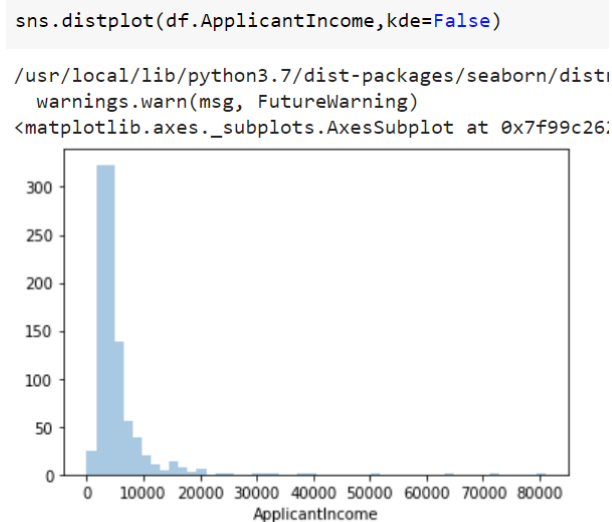
```
df['Loan_Status'].value_counts().plot(kind='pie', autopct='%1.2f%%', figsize=(6, 6))  
<matplotlib.axes._subplots.AxesSubplot at 0x7f99cd11d410>
```



The pie chart above represents the percentage of the loan status. Results show that there are more approved loans than disapproved.



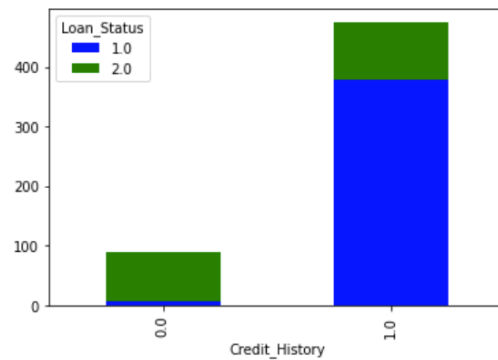
These histograms display the gender and marriage in accordance to the applicant income. It can be noticed that males have the highest income according to the data. Males that are married have greater income than unmarried male. And the same goes for females.



The histogram represents that people with better education should normally have a higher income.

```
temp3 = pd.crosstab(df['Credit_History'], df['Loan_Status'])  
temp3.plot(kind='bar', stacked=True, color=['blue', 'green'], grid=False)
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f99c23e5490>



This diagram outlines that the chances of getting a loan are higher if the applicant has a valid credit history.

**Summary:**