

Gramener Case Study

Sham Kumar

Naveen Shivashankar

Richa Pal

Ravi Babu Sinha

Gramener Case Study Overview

- **Problem Statement:**

A Consumer finance company which specialises in lending loans to urban customers has to make a decision for loan approval based on applicant's profile. There are two types of risks associated with it:

- * Not approving the loans for the applicant who are likely to repay the loan is a loss of business to the company
- * Approving the loan to the applicant's who are not likely to repay, is a financial loss to the company

- **Business Objective:**

To identify the patterns which indicates if a person is likely to default, which can be used to take actions such as denying loans, reducing the loan amount or lending at a higher interest rates.

- **Data Source:**

Masked data set from IIIT-B, which contains 39,717 rows and 111 columns

- **Approach:**

- Data collection for the analysis
- Review the data set and understanding the metrics
- Performed the EDA for the analysis
- Create segments for "Charged Off" and "Fully Paid" status categories
- Perform various analysis to find the pattern of the "Charged Off" segments
- Create various charts to visualize the pattern and present the findings

Gramener Case Study - Exploratory Data Analysis

- **Data Cleaning/Manipulation:**

- Checked for “Null” values in the data set. We had about 54 Columns which has no values at all, hence we have dropped them
- Dropped the columns with more than 50% of Null values in it.
- After reviewing the “Data Dictionary”, we had dropped 18 columns that are not relevant for the analysis
- We had imputed few columns which had null values with “NA” and 0s
 - Emp Length and Title was imputed with “NA”
 - Revol_Util was imputed with 0
- Data type of few metrics was changed as per the requirement

- **Univariate Analysis:**

- Performed the “Univariate Analysis” for the below metrics:
 - Analysed & Identified the % share of various “Loan Status” – Charged Off – 14%, Fully Paid – 83% and Current – 3%
 - Analysed “Emp_Length” and “Title” metrics for blanks and imputed “NA”
 - Replaced blanks with 0s for “Revol Util” metrics
 - Analysed the “Share” column and understood the % share of the grades from “Charged Off” loans
 - Analysed the “Purpose” in the loan dataset – “Debt Consolidation”, Credit Card” and “Other” purpose tops the list among all the “Purpose”
 - Analysed the “Home Ownership”, “Verification Status” and “Term” metrics for the “Default” loan status data set

- **Segmented Univariate Analysis:**

- Created different segments for “Charged Off” and “Fully Paid” Loan Status for effective analysis

- **Bivariate Analysis:**

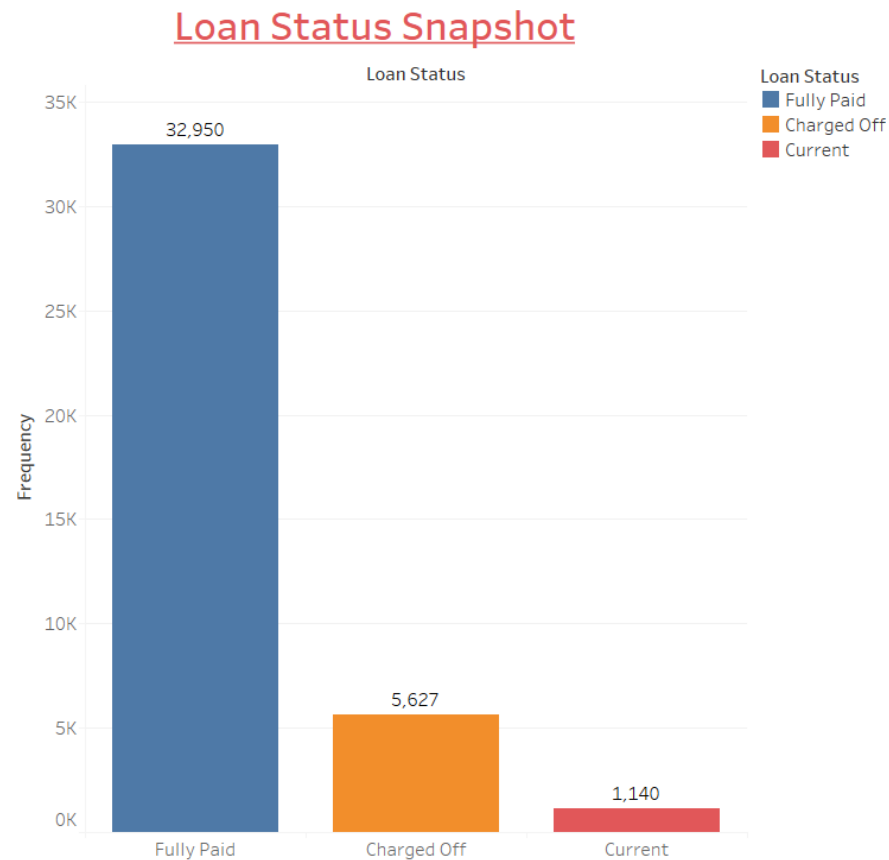
- Performed Bi-Variate analysis for the “Loan Status” metric with the below metrics:
 - Grade
 - Interest Rates
 - Income
 - Derived Income Segment
 - Pub_Rec
 - Experience
 - State
- Performed Bi-Variate analysis for the “Grade” with the “Annual Income” metrics:

- **Derived Metrics:**

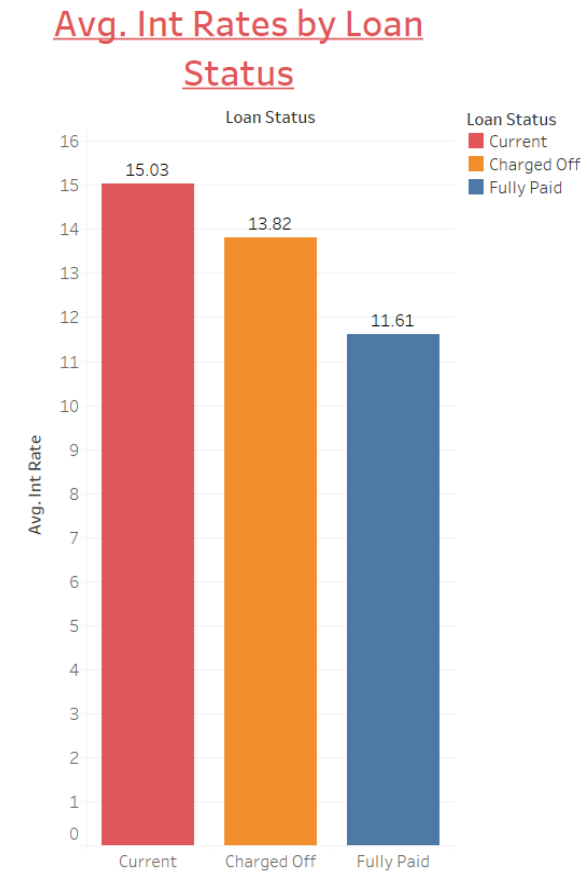
- Derived “Income Segments” from the “Annual Income” metric
 - Income \leq \$50K
 - Income between \$50K and \$100K
 - Income $>$ \$100K

Loan Status Snapshot

- There are 39,717 rows in the data set : 14% of them are from “Charged Off” status, 83% are from “Fully Paid” and 3% are from “Current” status
- Avg. interest rates for “Current” status is 15.03%, Charged Off is 13.82% and Fully Paid is 11.60%



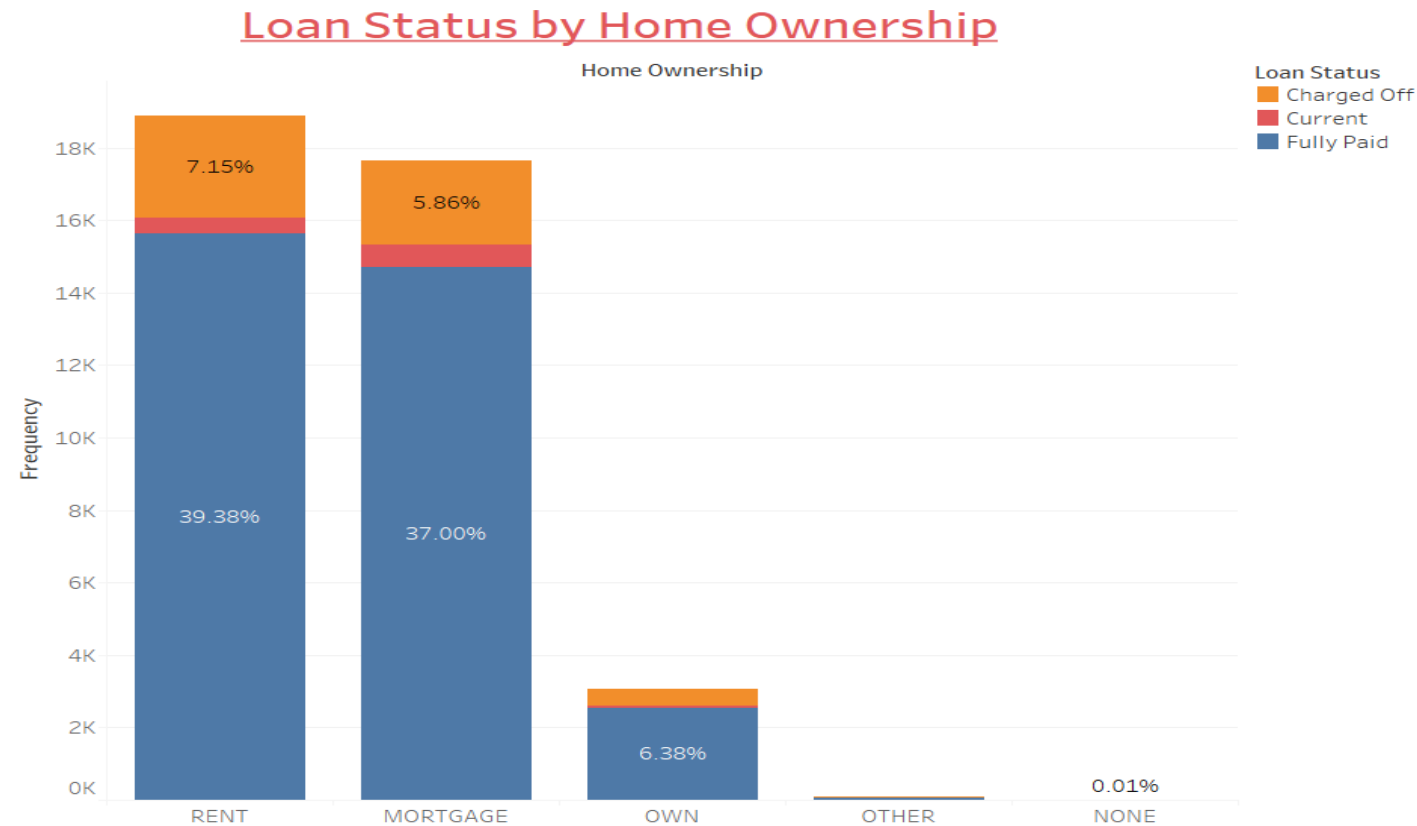
Count of Funded Amnt for each Loan Status. Color shows details about Loan Status. The marks are labeled by count of Funded Amnt.



Average of Int Rate for each Loan Status. Color shows details about Loan Status. The marks are labeled by average of Int Rate.

Loan Status Vs Home Ownership

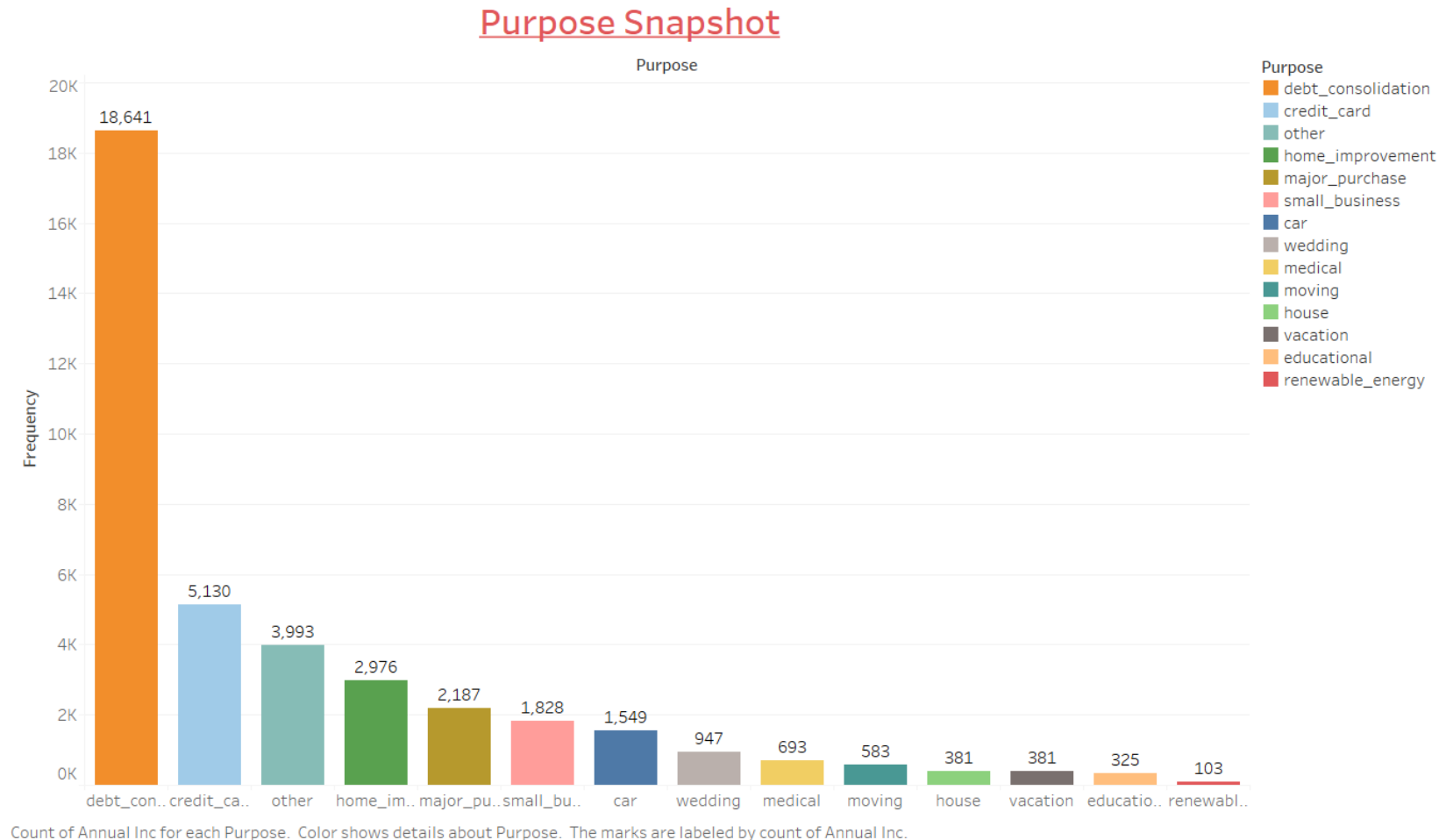
- Among the various “Home Ownership” types, majority of the borrowers stay on “Rent” followed by “Mortgage”
- “Charged Off” loan status also has the similar trend among the “Home Ownership”
 - 50% of them stay on “Rent”
 - 41% of them already have “Mortgage” loans
 - 8% have “Own” house



Count of Annual Inc for each Home Ownership. Color shows details about Loan Status. The marks are labeled by % of Total Count of Annual Inc.

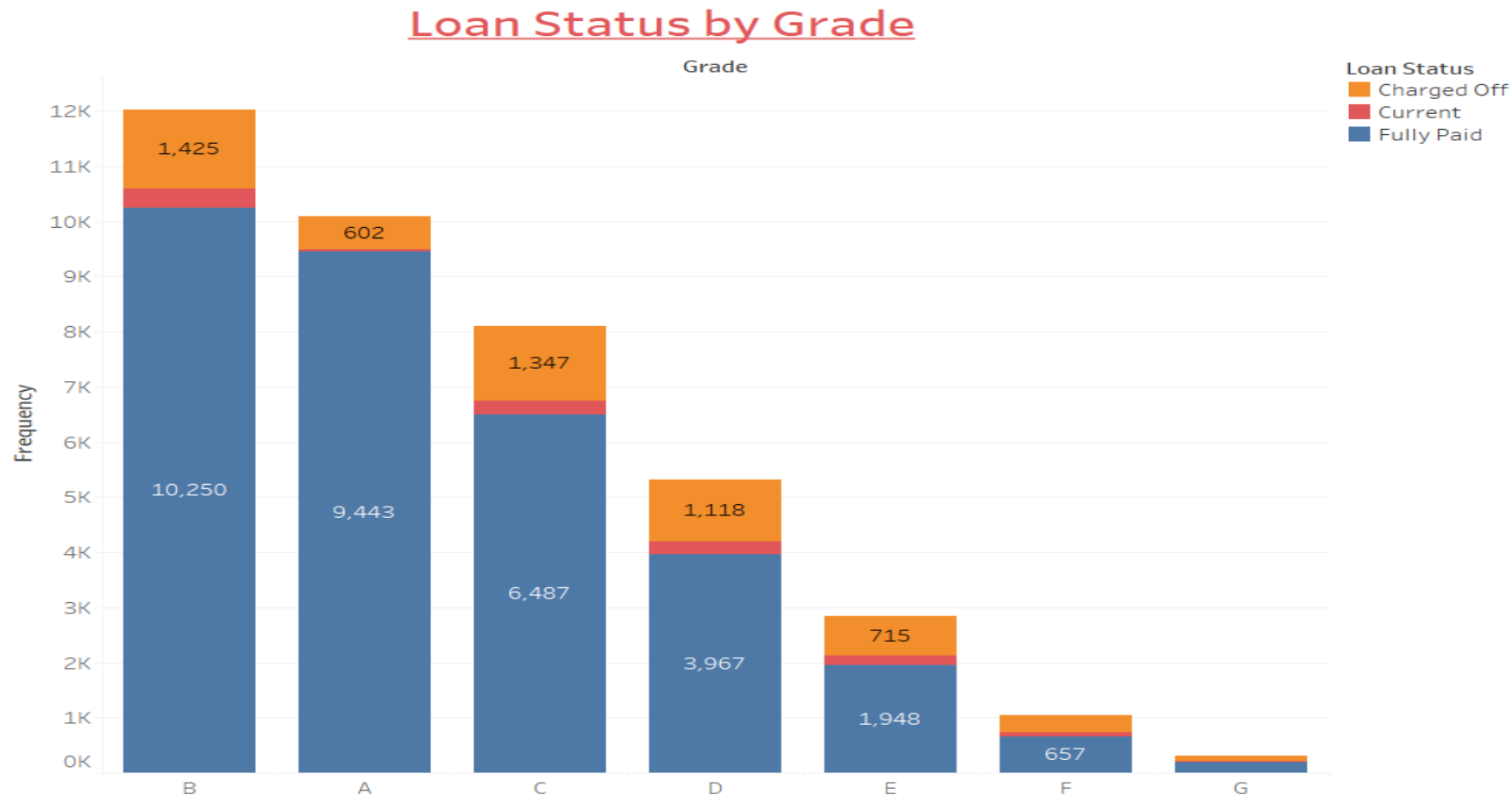
Loan Purpose Analysis

- Among the various “Purpose” of loans, majority of them are for “Debt Consolidation” followed by “Credit Card” and “Other” purposes
 - 47% are for “Debt Consolidation”, 13 % are for “Credit Card” and 10% are for “Other” purposes



Loan Status by Grade

- Majority of the loans are availed by “B” grade category. This grade also has the highest “Full Paid” rates compared to other grades
 - B, C & D grade category has the highest number of “Defaulters” whereas
 - A, F & G grade categories has the lowest number of “Defaulters”

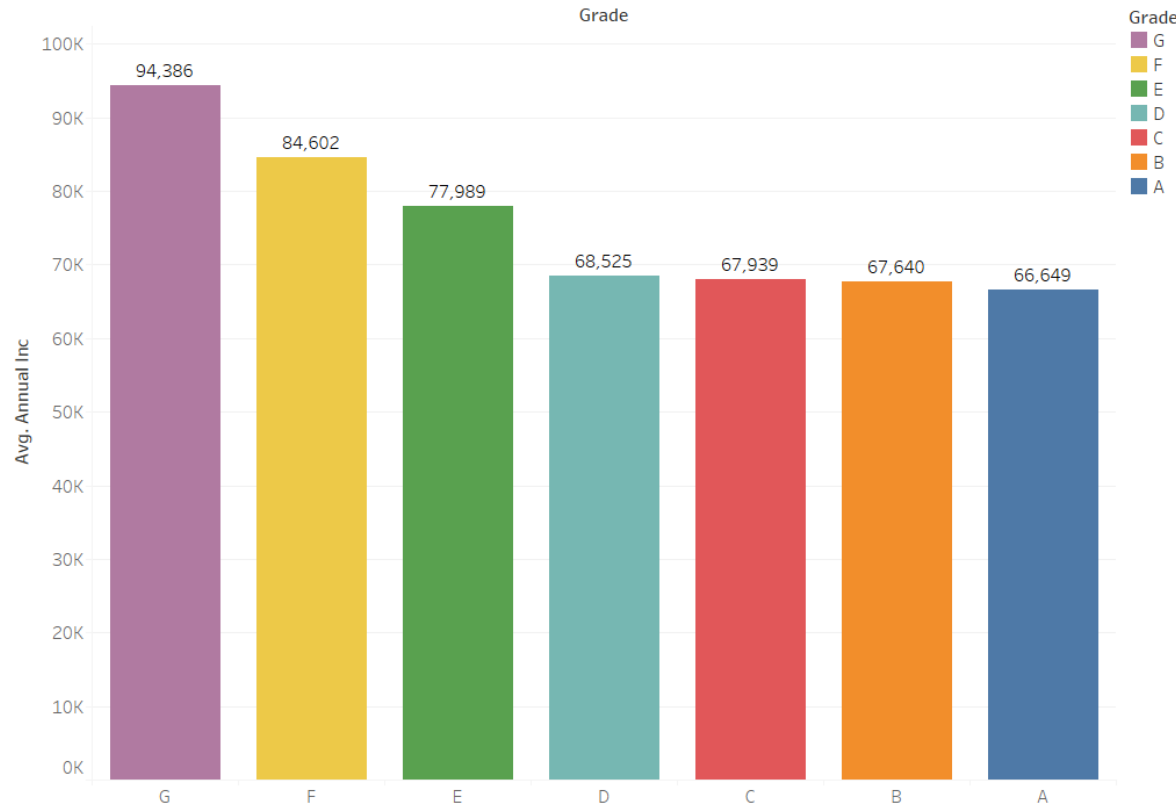


Count of Annual Inc for each Grade. Color shows details about Loan Status. The marks are labeled by count of Annual Inc.

Average Income analysis

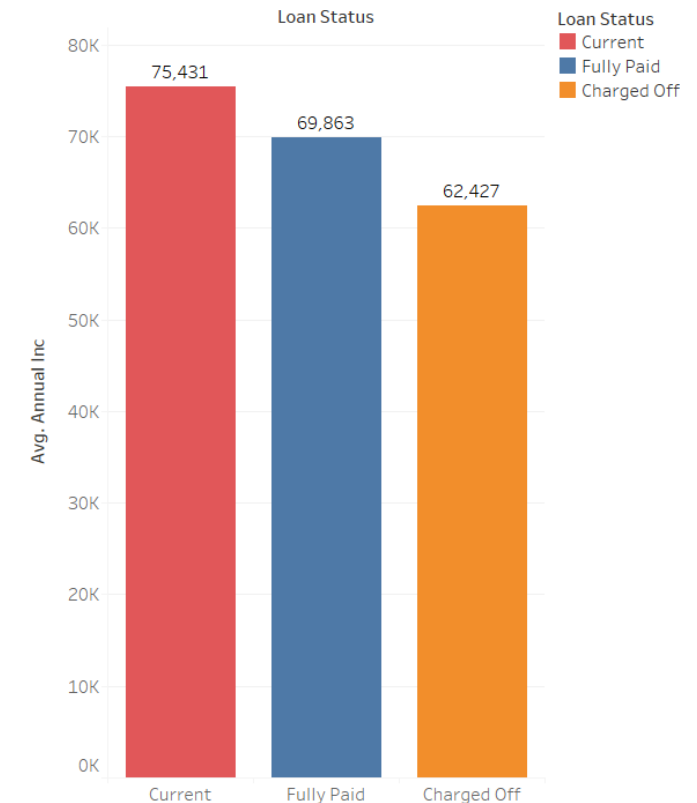
- G, F & E grade category has the highest average income among all the grades, hence these grades has the lowest “Default” rates
 - Average income for A, B, C & D grades are almost inline and avail the most number of loans Vs other 3 grades
 - Average income by Loan status : Current - \$75.4K, Fully Paid - \$69.8K and Charged Off - \$62.4K

Avg. Income by Grade



Average of Annual Inc for each Grade. Color shows details about Grade. The marks are labeled by average of Annual Inc.

Avg. Income by Loan Status

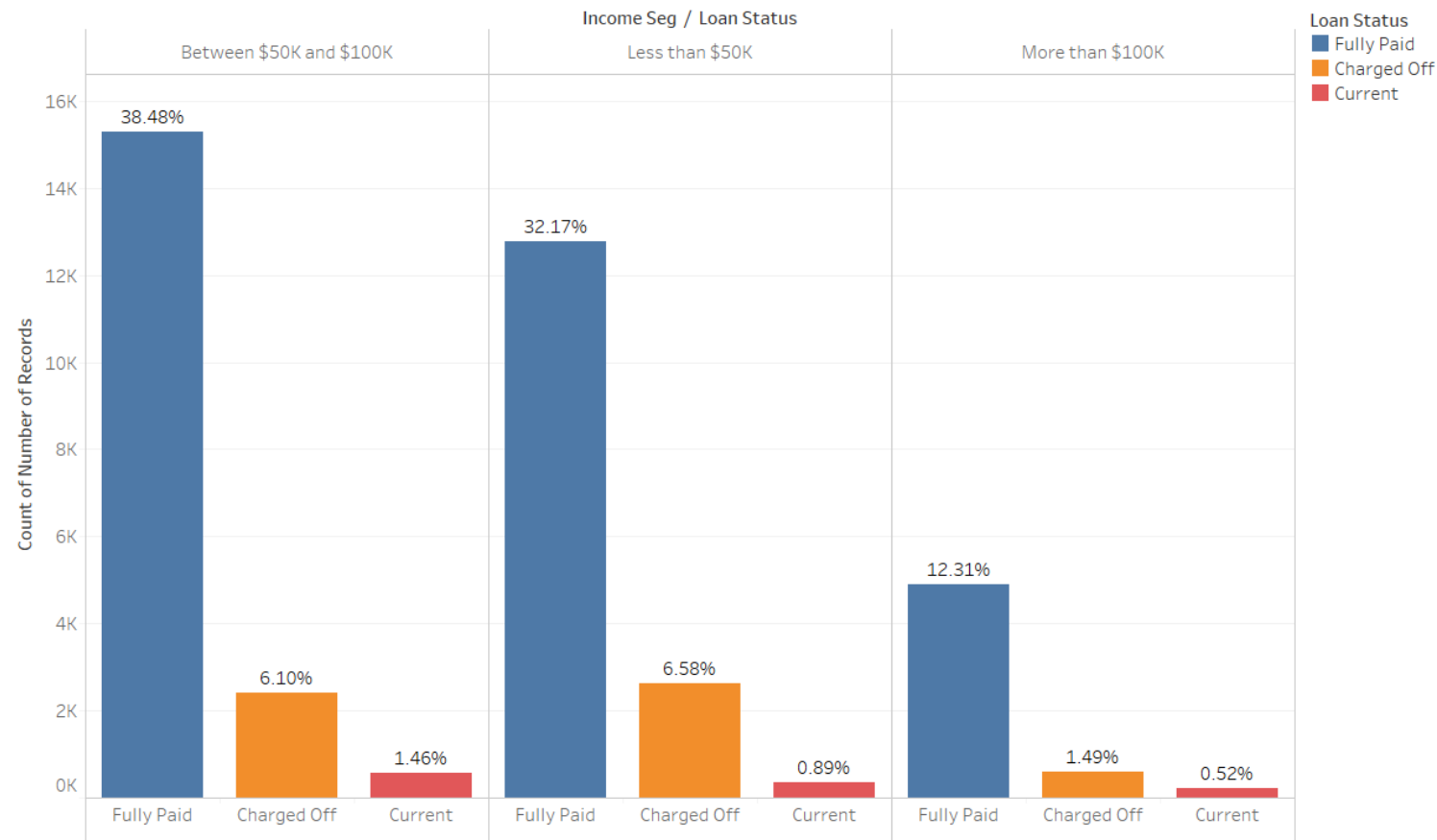


Average of Annual Inc for each Loan Status. Color shows details about Loan Status. The marks are labeled by average of Annual Inc.

Loan Status by Income Segment

- Applicant's with income range between \$50K and \$100K have the highest "Full Paid" rates
- Among the "Defaulter's" segment, applicant's with income range less than \$50K tops the list
- Income range >\$100K avails less loans and hence they have the least "Full Paid" and "Charged Off" rates

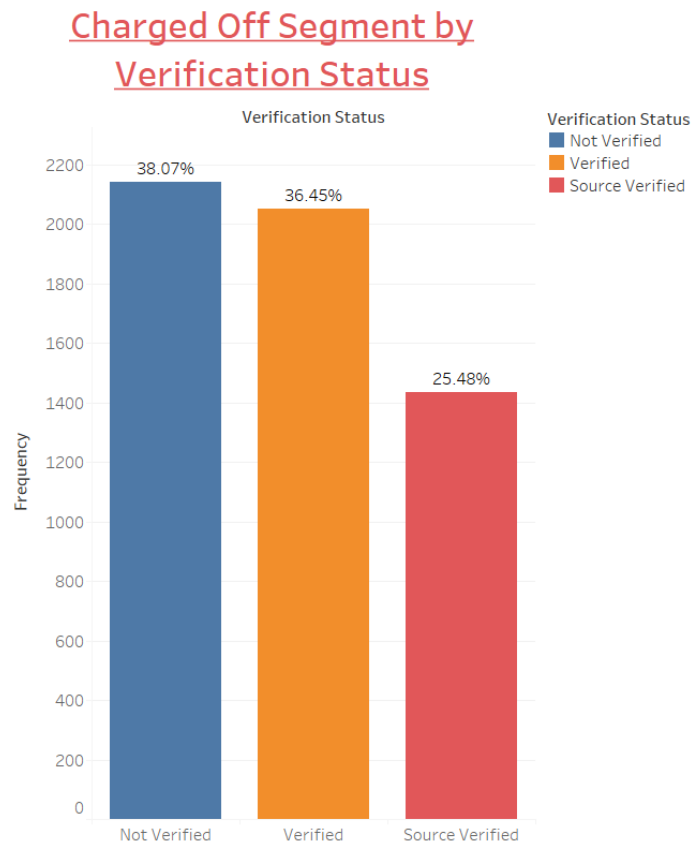
Income Segment by Loan Status



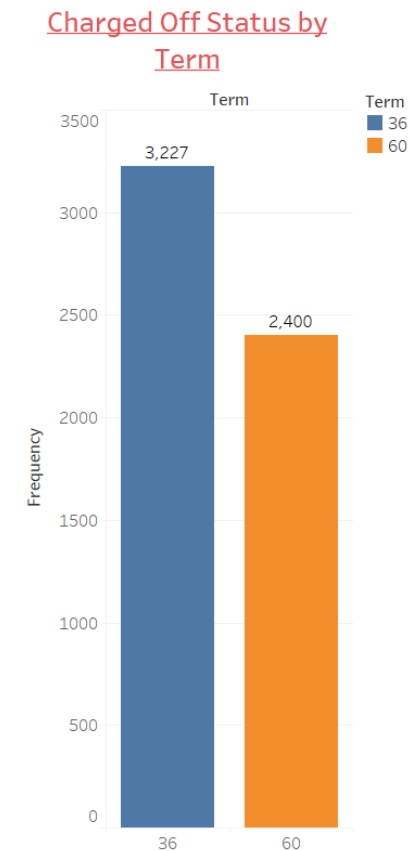
Count of Number of Records for each Loan Status broken down by Income Seg. Color shows details about Loan Status. The marks are labeled by % of Total Count of Number of Records.

Charged Off status by Verification status and Term

- 38% of the “Charged Off” applicant’s application was not verified while approving the loan, hence this segment has the highest number of “Defaulter’s” rate
- Among the “Defaulter’s” segment, majority of the applicant’s opt for 36 month’s term Vs 60 month’s



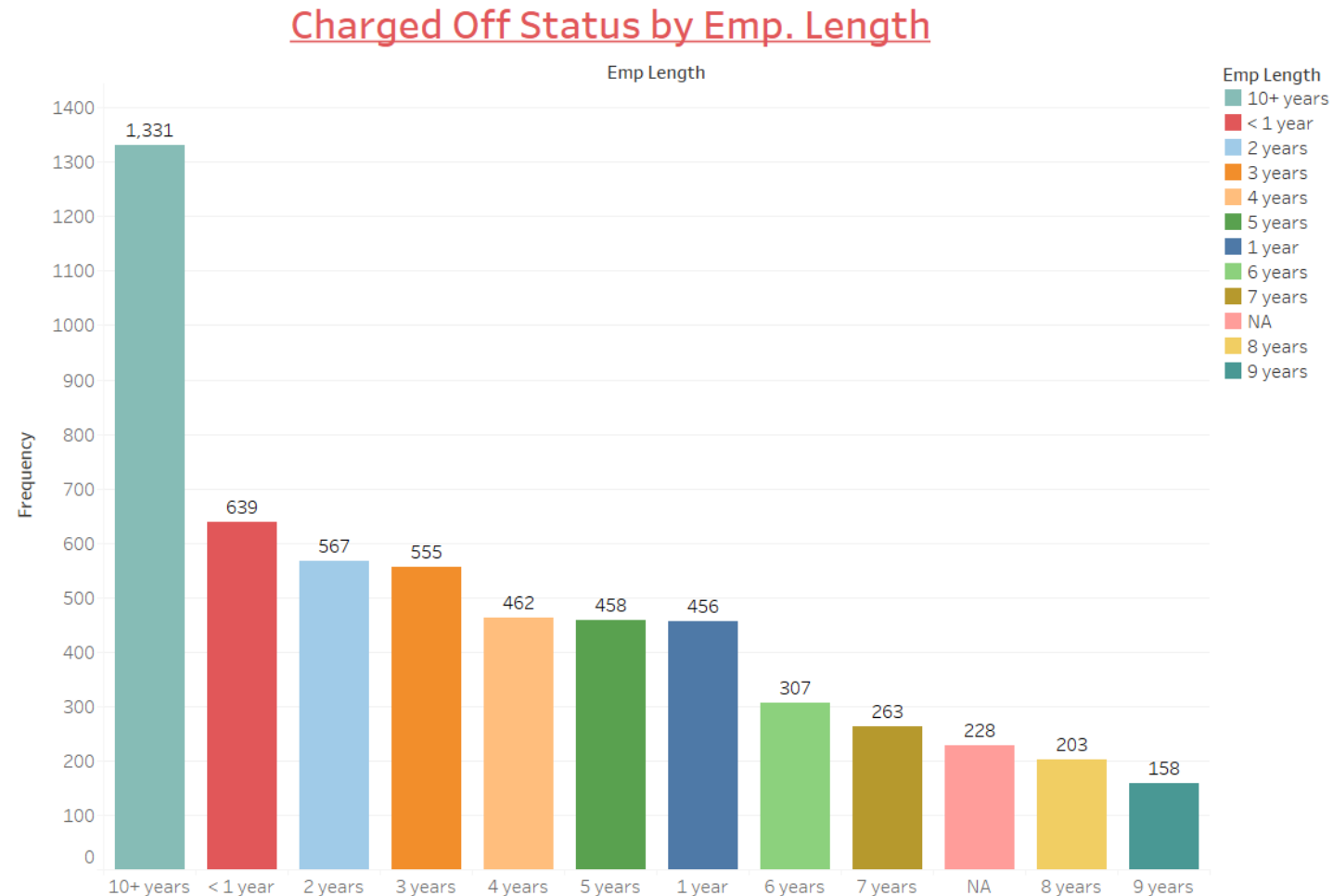
Count of Number of Records for each Verification Status. Color shows details about Verification Status. The marks are labeled by % of Total Count of Number of Records. The data is filtered on Loan Status, which keeps Charged Off.



Count of Number of Records for each Term. Color shows details about Term. The marks are labeled by count of Number of Records. The data is filtered on Loan Status, which keeps Charged Off.

Charged Off status by Employee Length

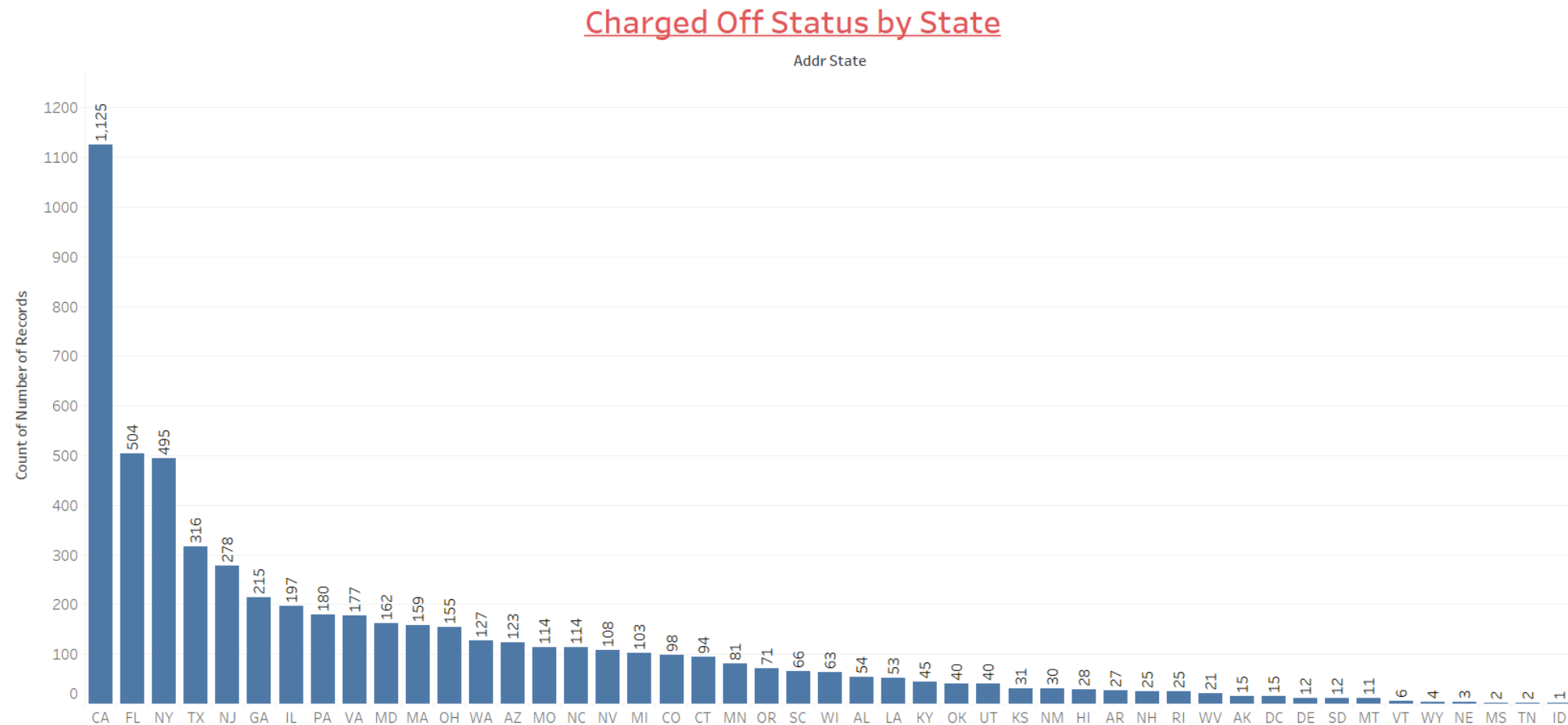
- Majority of the applicant's who have defaulted, have 10+ years of experience
 - 23.65% of the applicant's have 10+ years of work experience



Count of Number of Records for each Emp Length. Color shows details about Emp Length. The marks are labeled by count of Number of Records. The data is filtered on Loan Status, which keeps Charged Off.

Charged Off status by State

- California state has the maximum number of “Defaulter’s” followed by “Florida” and “New York”
 - 20% are from California and 9% are from Florida and New York



Count of Number of Records for each Addr State. The marks are labeled by count of Number of Records. The data is filtered on Loan Status, which keeps Charged Off.

- “Charged Off” segments comprises of 14% of the total dataset
- Loans that are “Charged Off” are primarily taken for “Debt Consolidation” (47%) followed by “Credit Card” (13%)
- Loans with grade B, C & D has the highest number of “Defaulter” rate Vs other grades
- Avg. Interest rates for the “Charged Off” loans is 13.82%
- Avg. Income of applicant’s who have defaulted is around \$62.4K
- Most number of derogatory public records are for the “Defaulter” segments
- Defaulter’s segment mostly stay on rent or they have a “Mortgage” loans
 - 51% stay on Rent, 41% have a “Mortgage” loans and only 8% own a home
- Most of the application of the “Charged Off” loans has not been verified properly, hence they have the highest default rate
 - 38% of the application for the loans that were “Charged Off”, were not verified
- Number of “Defaulter’s” are more for the people who have availed for 36 month’s term
 - 57% have opted for 36 month’s term
- Majority of the applicant’s who have defaulted, have 10+ years of experience
 - 23.65% have an experience of 10+ years
- California state has the maximum number of “Defaulter’s” followed by “Florida” and “New York”
 - 20% are from California and 9% are from Florida and New York

Thank You