

Lending Club Case Study

Exploratory Data analysis

By Tamal Saha & Dharmendra Singh Chouhan

Executive Summary

Data Background

- The data contains information about past loan applicants and whether they 'defaulted' or not.
- The dataset contains information starting from 2007 to 2011

Objective

- The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

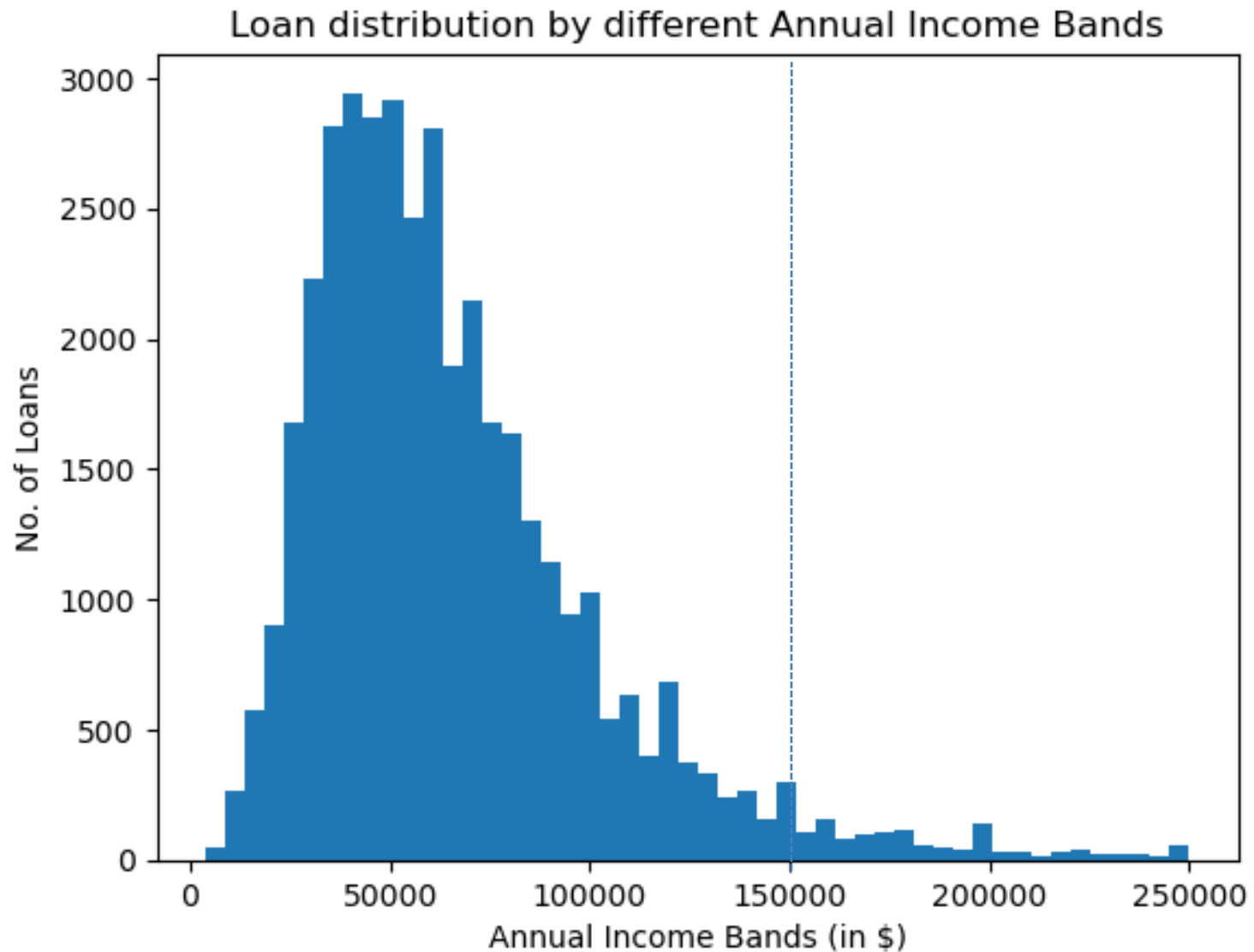
Analysis

- Data Cleaning – Removing Columns with 100% null values, presence of extreme values, reclassifying existing column types
- Univariate, Bi-variate analysis

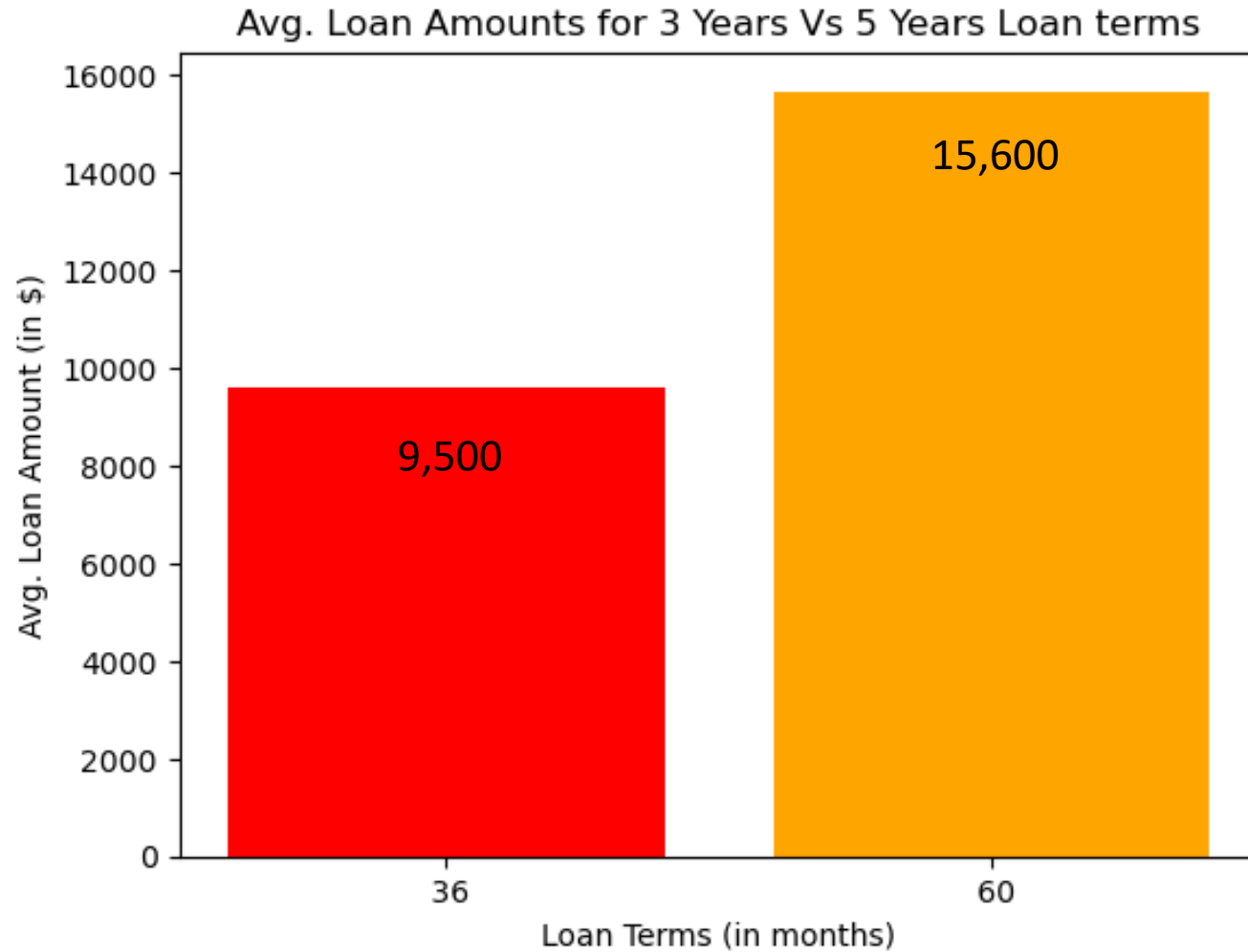
Key Observations

- ~14% Charge-off accounts
- Top 5 states (CA, FL, NY, TX and NJ) account for ~48% of the charge-off accounts
- Income, Grade are key driving factors to identify charge-off accounts

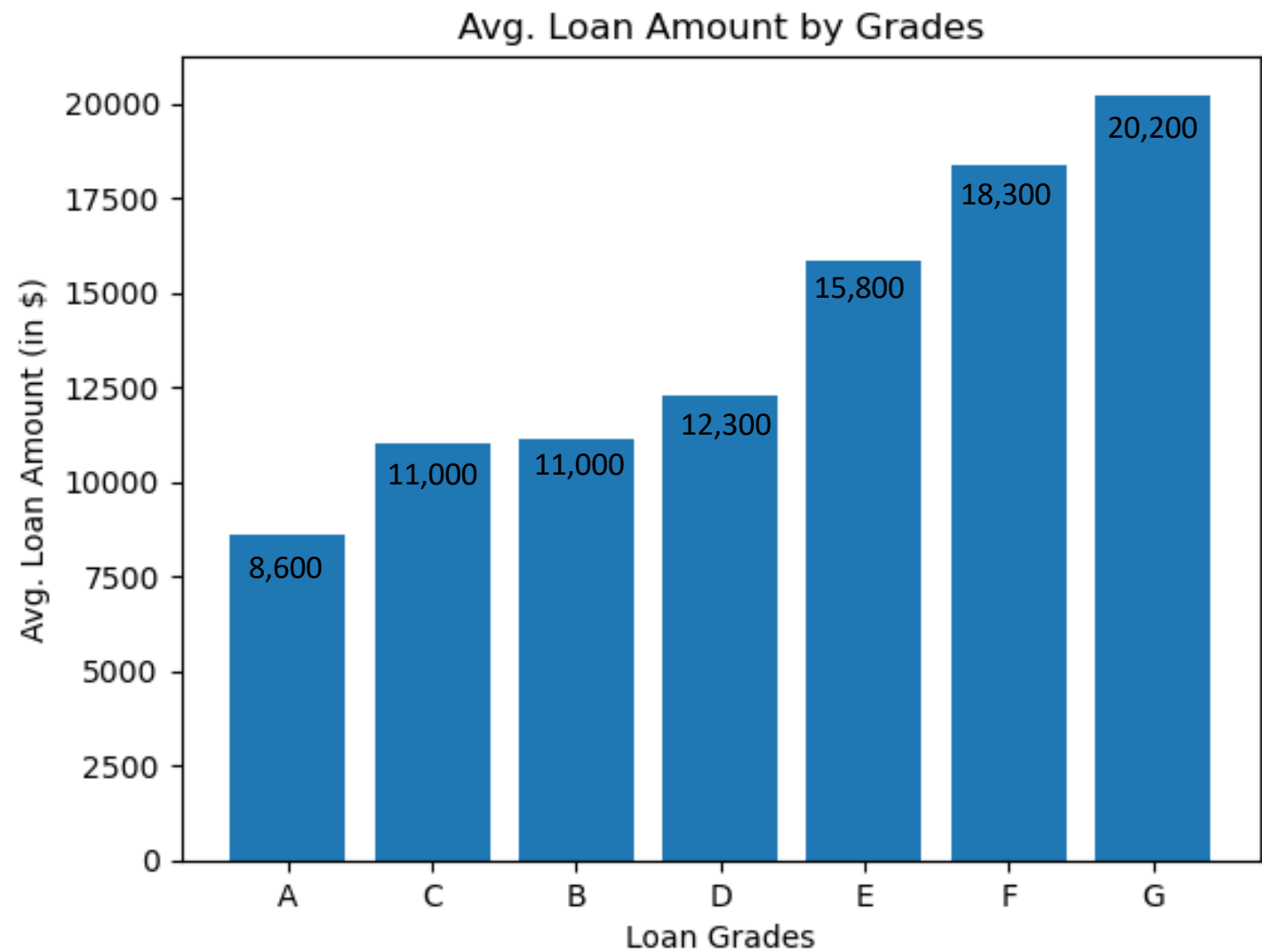
~95% of the loans are provided to the customers with annual income less than \$150k



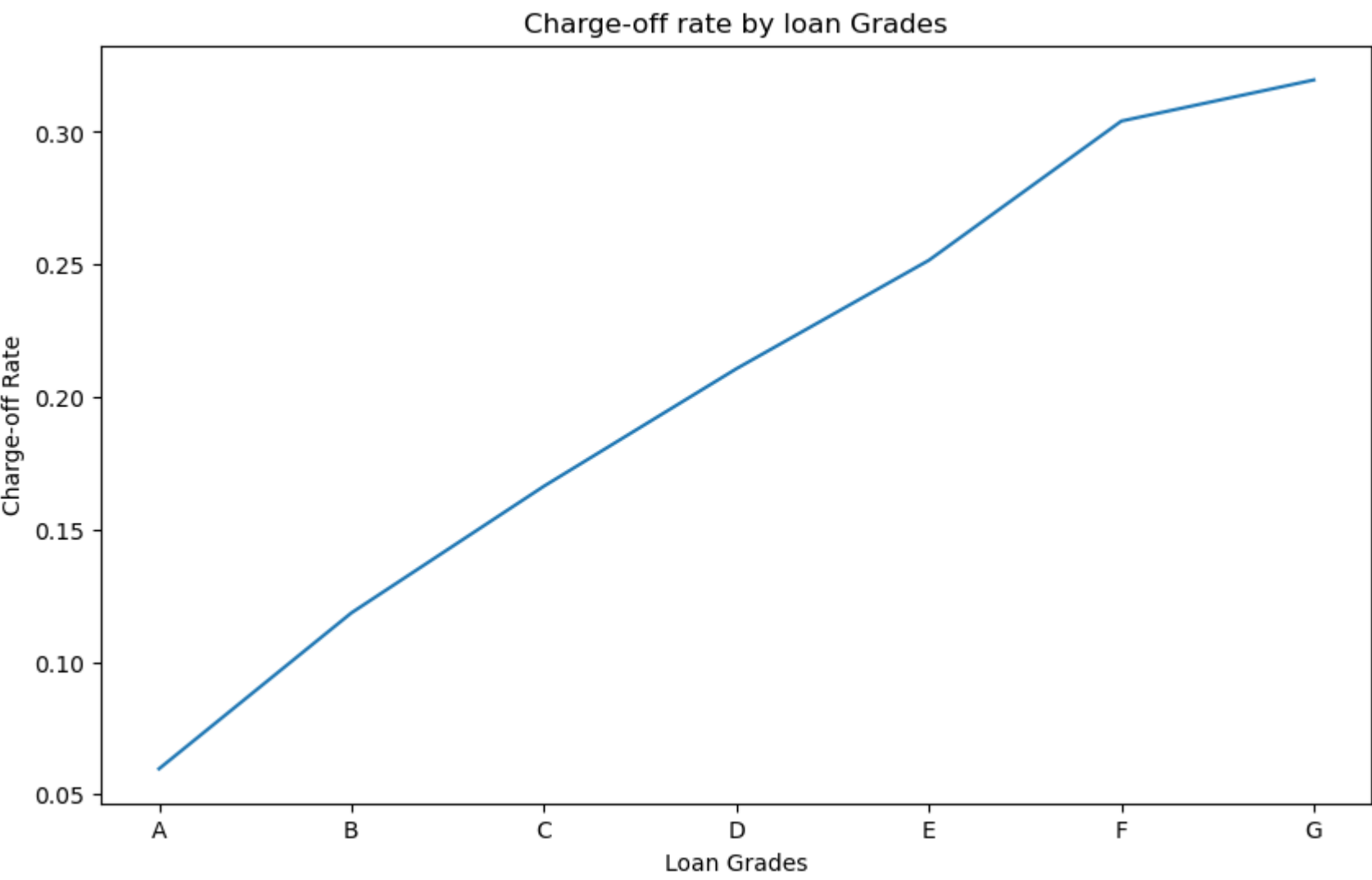
Overall Avg loan size is ~\$11.2k. Avg Loan amounts are ~60% higher for 5 years than 3 years tenure



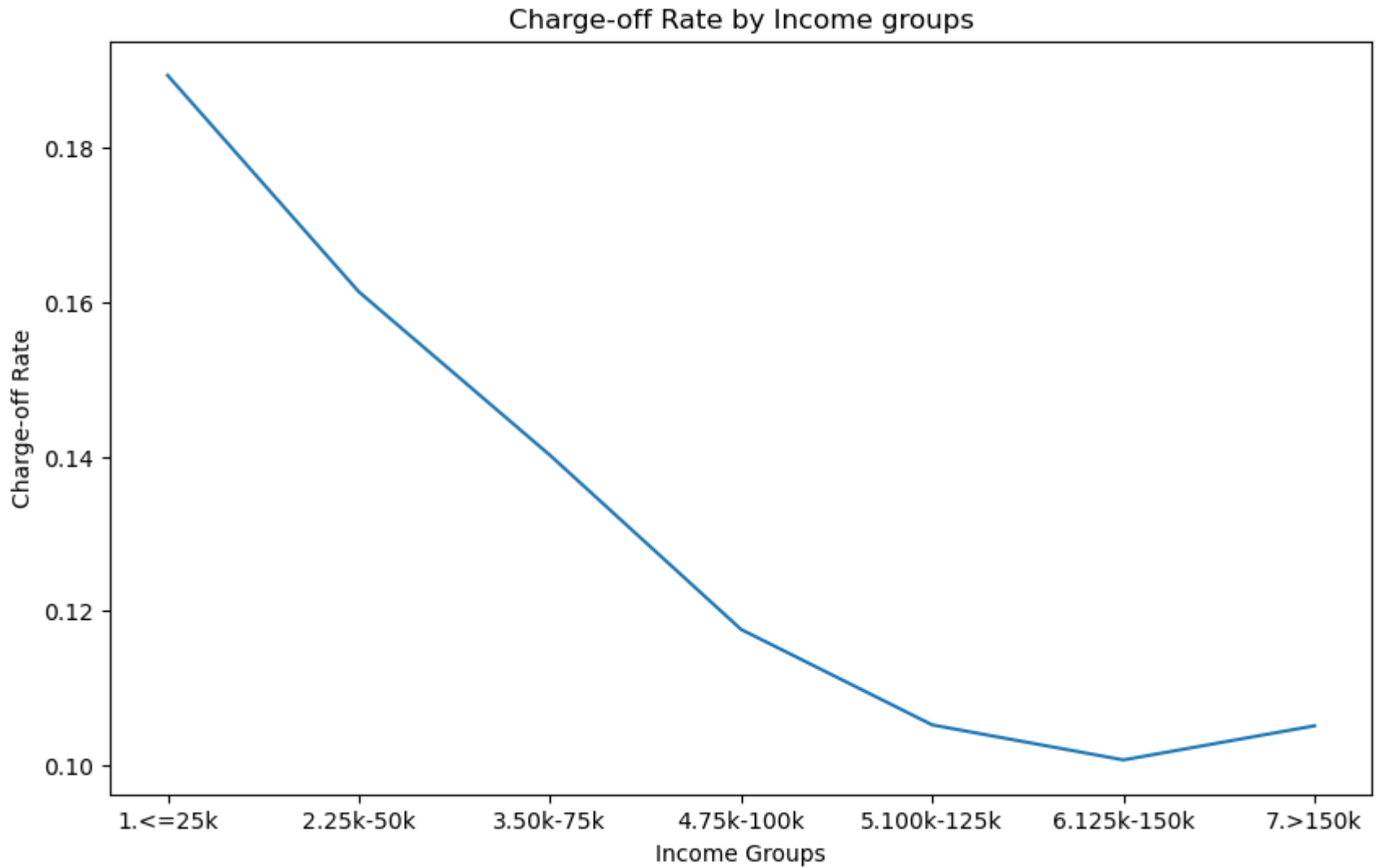
Avg Loan amounts increase from loan grade 'A' to 'G'



Increasing Charge-off rates from grade A to G shows the increasing risk and loan quality deterioration



Customers with higher annual income has less chance of getting charged-off



Charge-off accounts have higher revolving dollar percentage than other accounts

