Lending Club Case Study

Group Facilitator: Sanjay Saini

Group Member: Manish Ranjan

GitHub Repo URL: https://github.com/sanjaysaini-msn/LendingClubCaseStudy

- Business Requirement
- Approach
- Assumptions
- Some Insights
  - Univariate Analysis
  - Segmented Univariate Analysis
  - Bivariate Analysis
  - Probability Analysis
- Conclusion

Business Requirement

- Help the finance company in decision making for loan approval based on the applicant's profile

- Risks involved

  - Not approving loans for the applicants who are likely to repay the loan

  - Approving loans for the applicants who are likely to default

- Identify patterns using past loan application data that can trigger charge-off loans

Approach

- Data fields understanding
  - Understand the purpose of different fields as available in the dataset
- Data Pre-processing
  - Remove columns having all null values or greater than a threshold
  - Remove columns which are not required or have no relation to the business requirement
  - Remove rows with missing values using a threshold
  - Impute the missing values for columns having missing values less than the threshold
  - Convert data types for int, float and date fields
  - Create derived columns
  - Remove outliers
- Exploratory Data Analysis
  - Visualize data and Identify important variables for decision making

Assumptions

- Rows with Loan status "Current" are not eligible for this analysis
- Threshold taken for Data pre-processing
  - Columns having null values in 35% or above rows
  - Rows having null values in 10 or more columns
- Categorical variables that can trigger defaults:
  - Employment length
  - Grade/Sub-Grade
  - Loan Purpose
  - Location (State Code)
  - Loan Amount
  - Interest Rate
  - Annual Income

Insights – Univariate Analysis

Observation:

➡ 85.24% loans are "Fully Paid" where as 14.75% loans are "Charged Off"

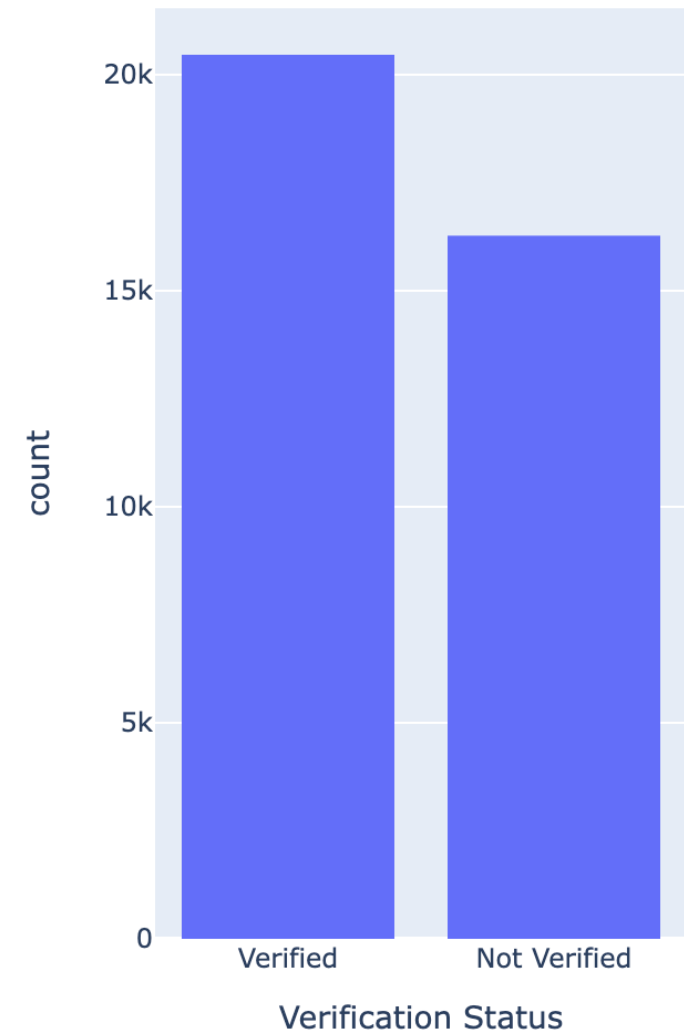Univariate analysis - Loan Status and Respective Percent
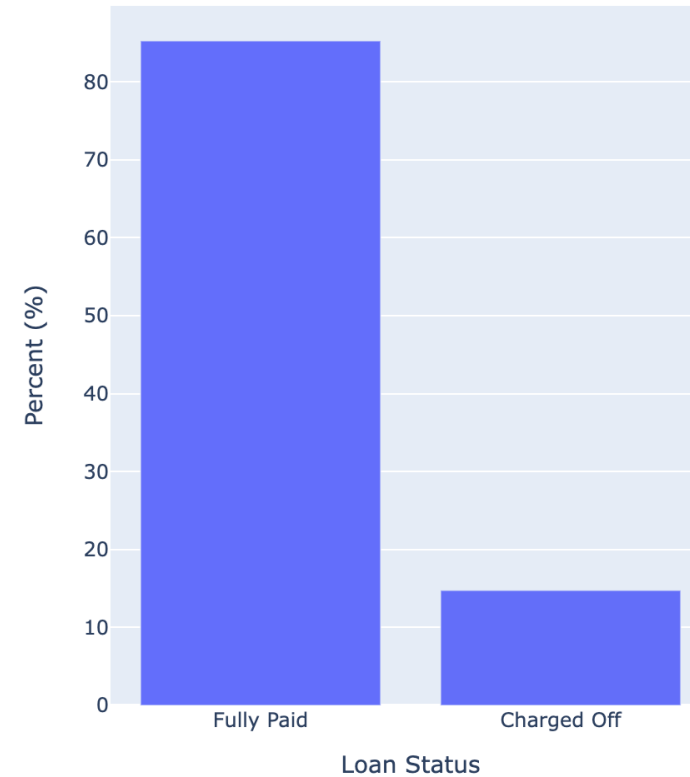
Insights – Univariate Analysis

Observation:

- Number of Non-Verified loans is quite high
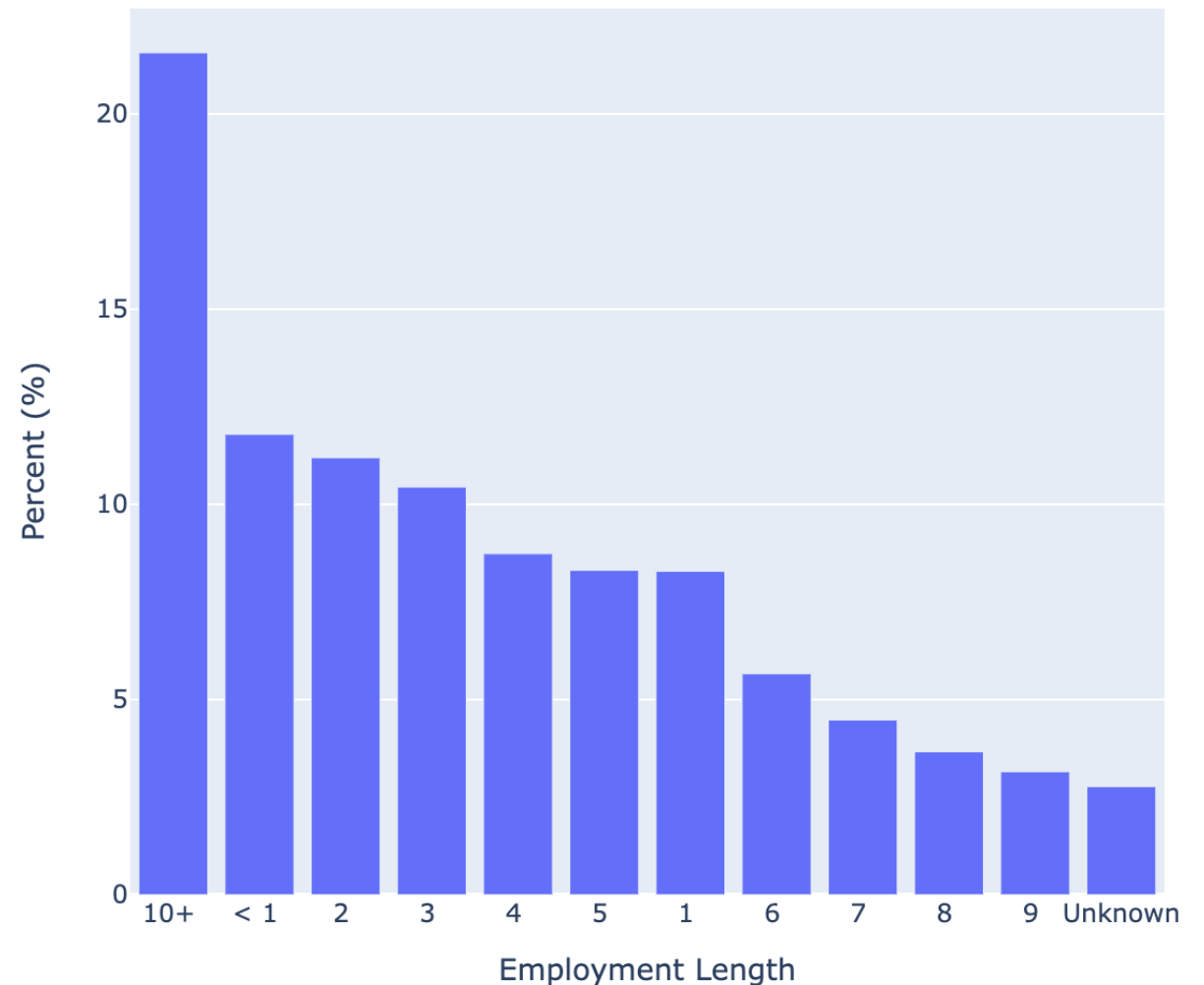

Univariate analysis for Verification Status

Insights – Univariate Analysis

Observation:

➤ More than 21% loans were given to the applicants having 10+ years of employment

Univariate analysis - Employment Length and Respective Percent
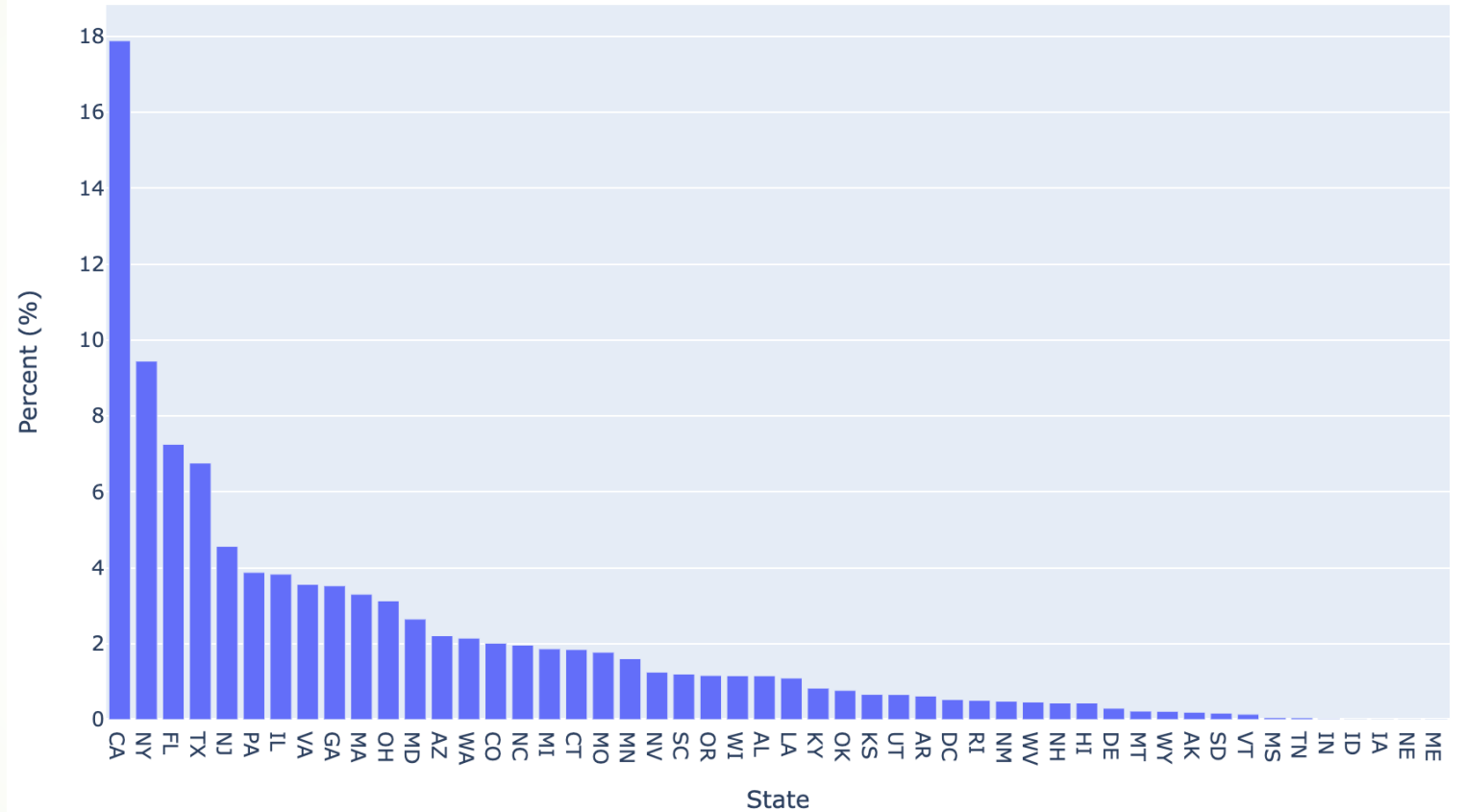
Insights – Univariate Analysis

Observation:

- Around 18% loans were given to applicants from California
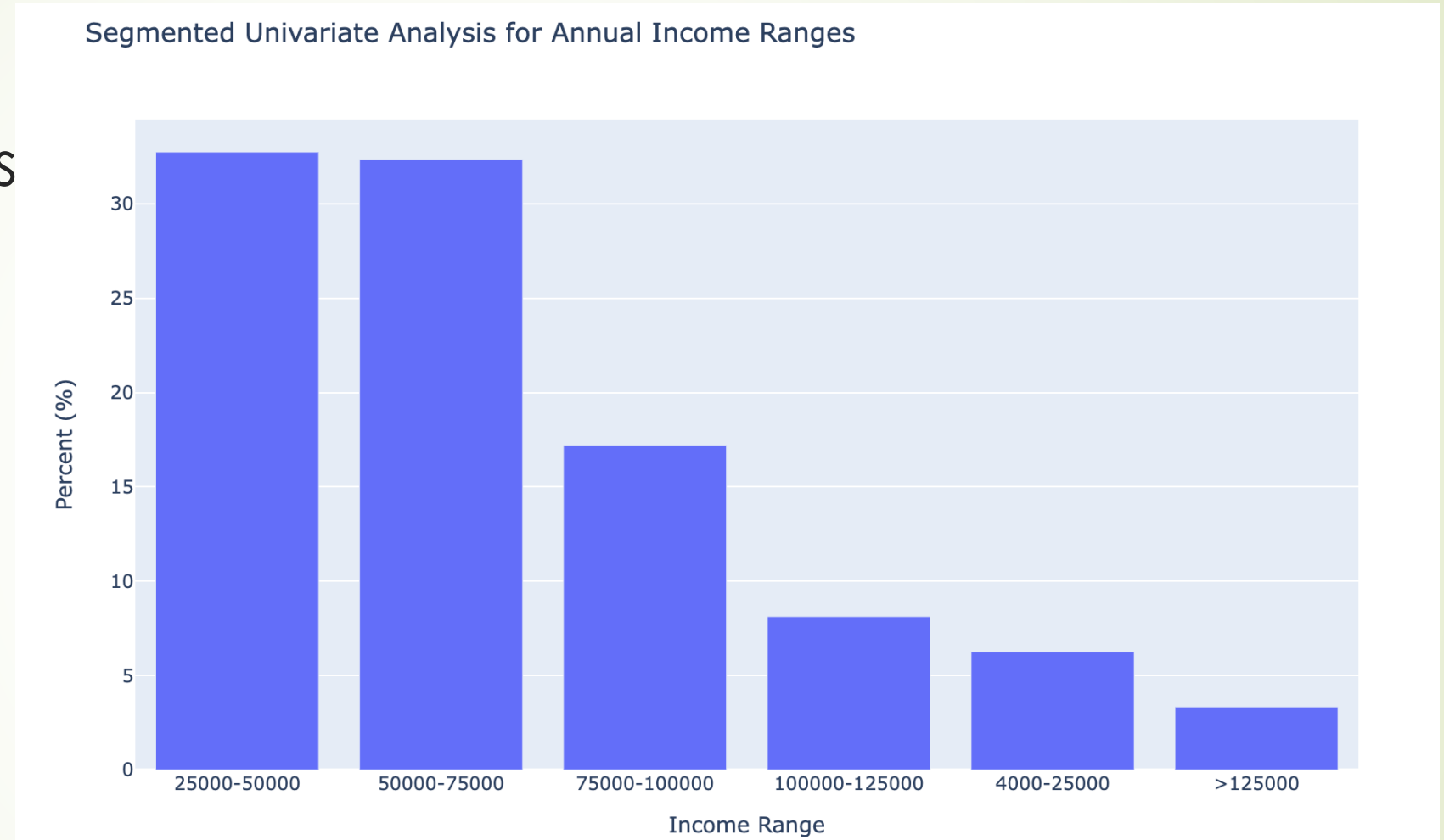

Univariate analysis - Location and Respective Percent

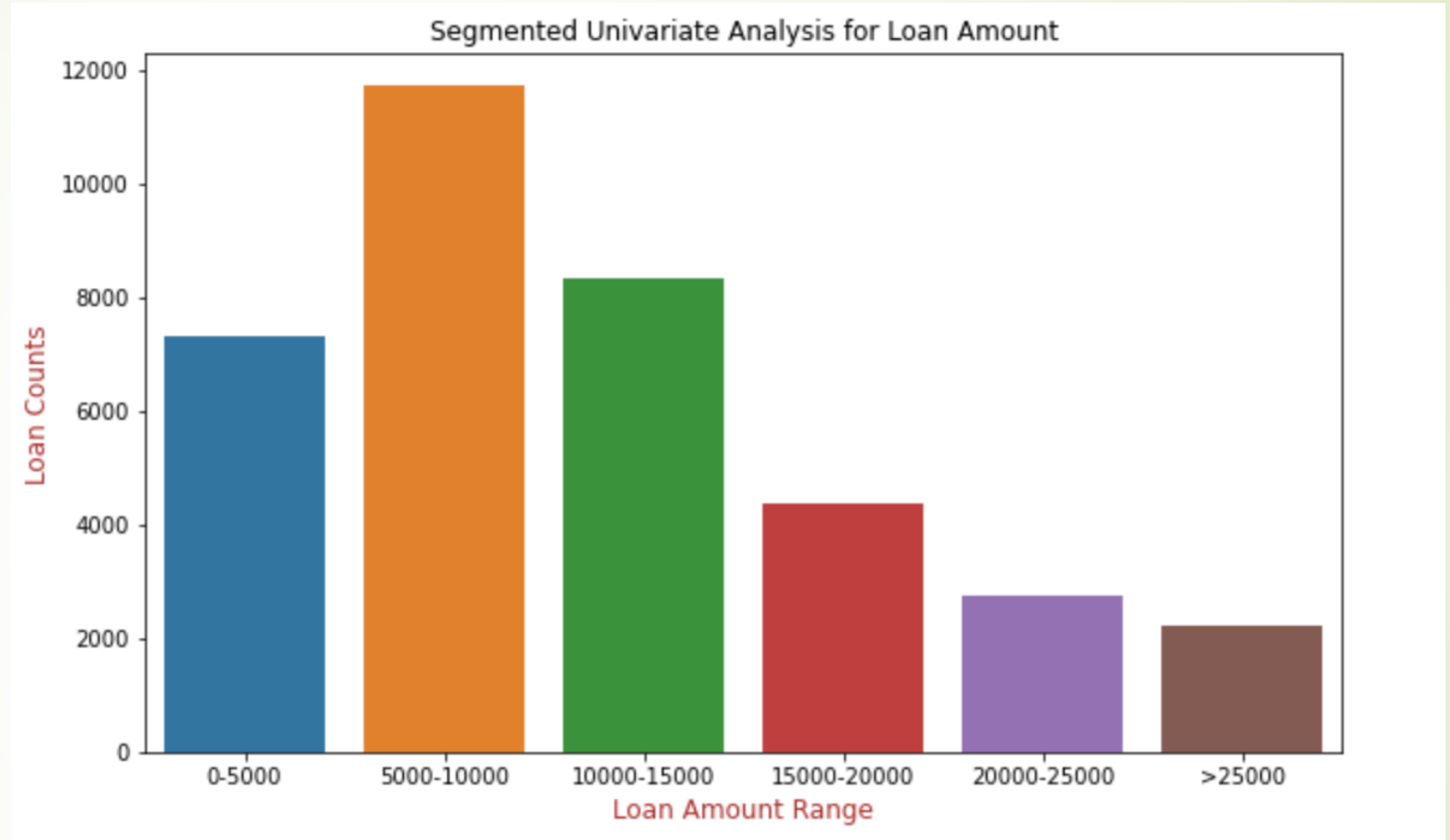# Insights – Segmented Univariate Analysis

Observation:

- Annual income of 64% applicants is between 25k and 75k

Segmented Univariate Analysis for Annual Income Ranges
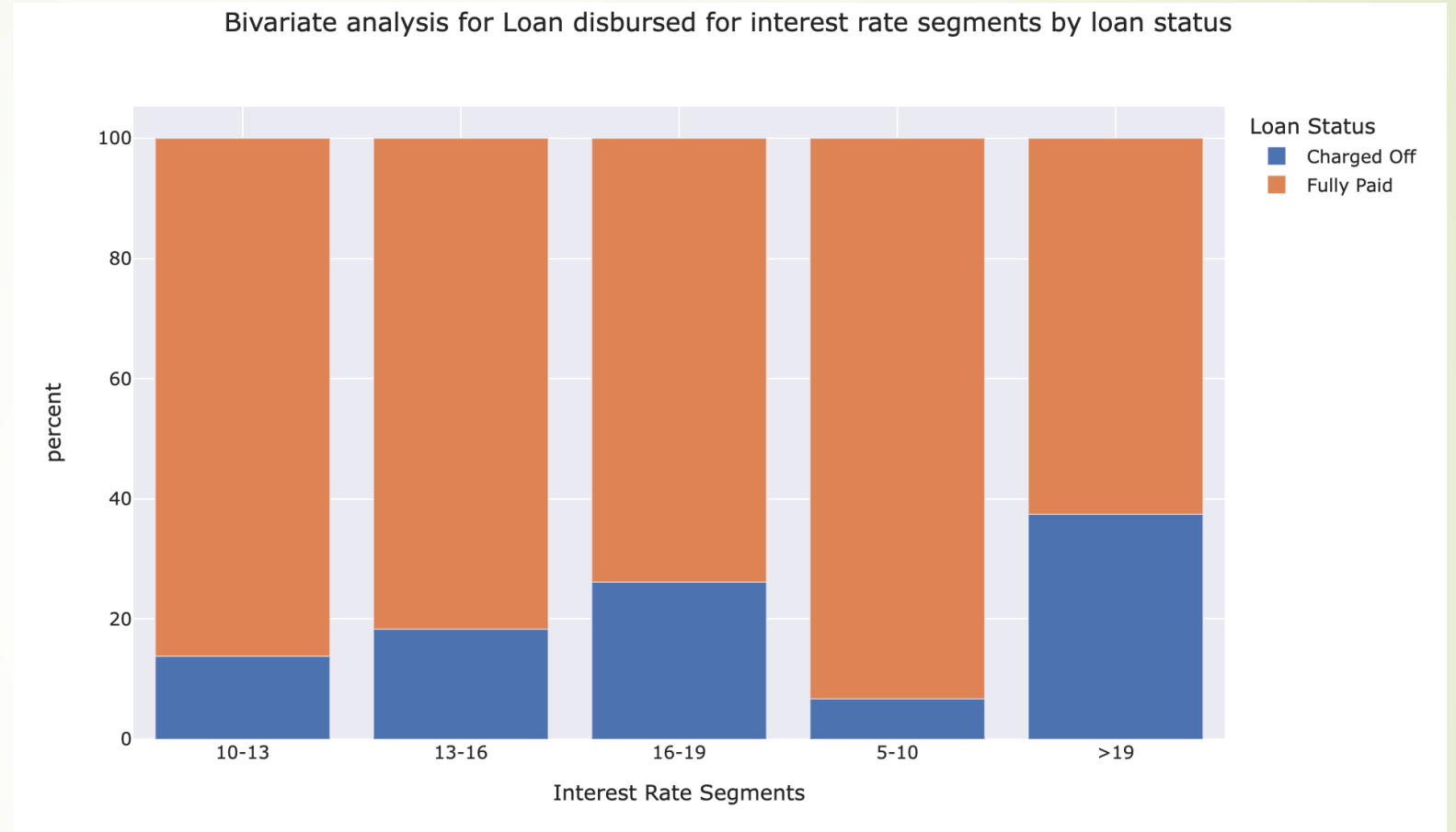
Insights – Segmented Univariate Analysis

Observation:

- Maximum loans disbursed are between 5k and 10k loan amount range

Segmented Univariate Analysis for Loan Amount
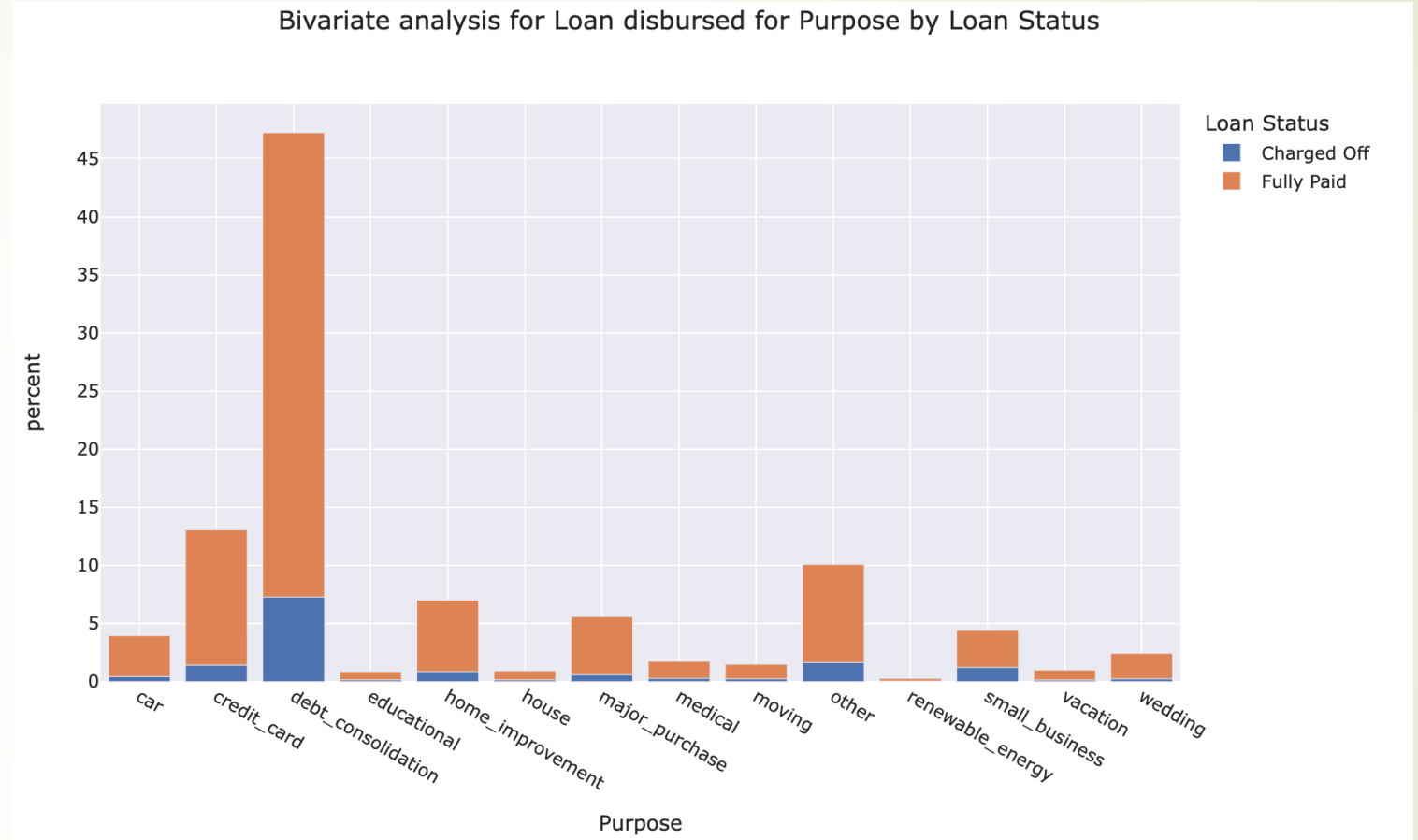
Insights – Bivariate Analysis

Observation:

- Defaulters % is low for lower interest rate (5-13%)

Bivariate analysis for Loan disbursed for interest rate segments by loan status
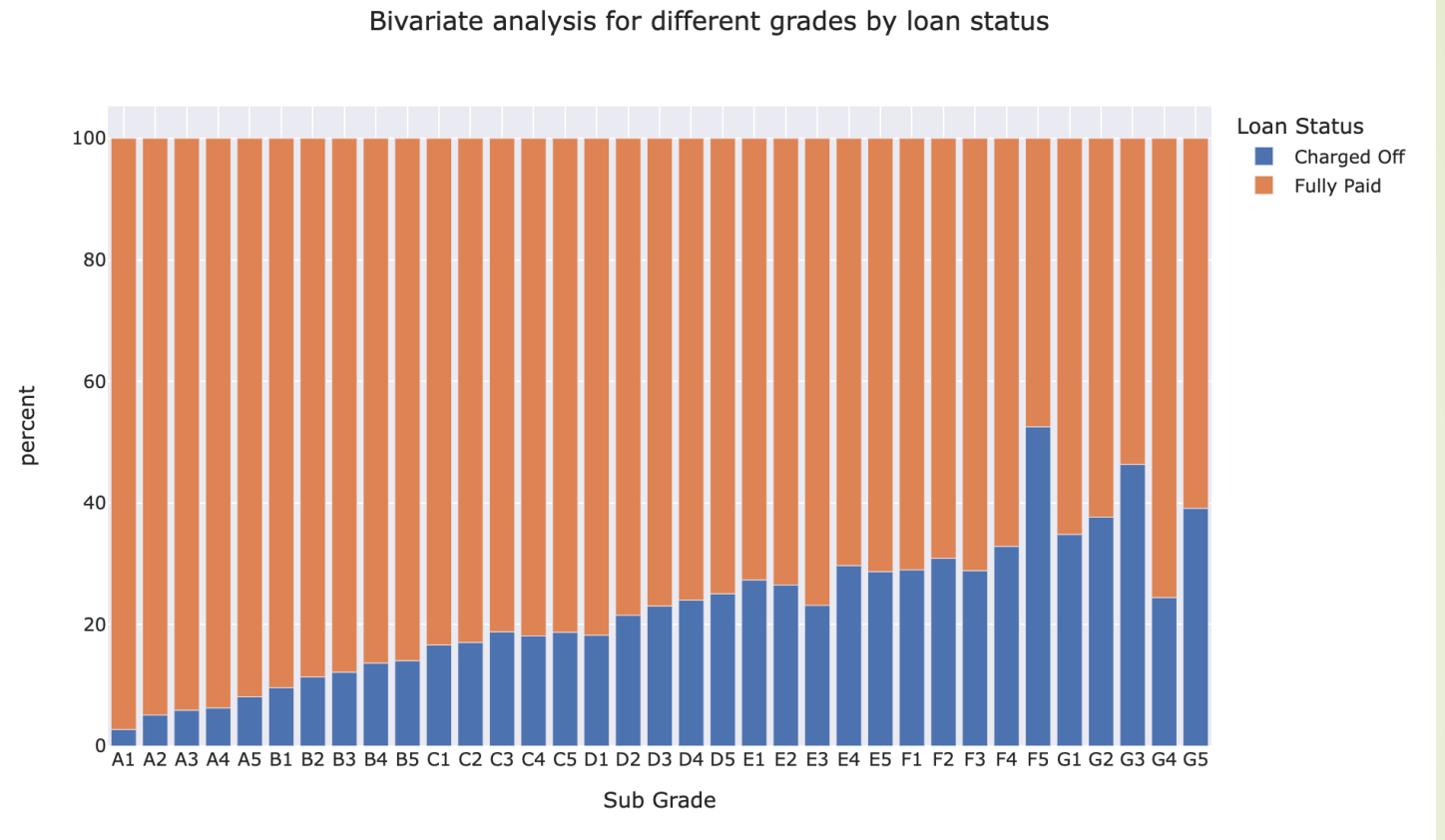
Insights – Bivariate Analysis

Observation:

➡ Applicants who have taken loan for debt consolidation are having highest % of charged off


Bivariate analysis for Loan disbursed for Purpose by Loan Status

Insights – Bivariate Analysis

Observation:

➡ Charged off % is low for grade A

➡ Charged off % is high for sub-grade F5

Bivariate analysis for different grades by loan status

Chart legend — Loan Status:
- Charged Off
- Fully Paid

Y-axis: percent (0 to 100)

X-axis (Sub Grade): A1 A2 A3 A4 A5 B1 B2 B3 B4 B5 C1 C2 C3 C4 C5 D1 D2 D3 D4 D5 E1 E2 E3 E4 E5 F1 F2 F3 F4 F5 G1 G2 G3 G4 G5

Insights – Probability Analysis
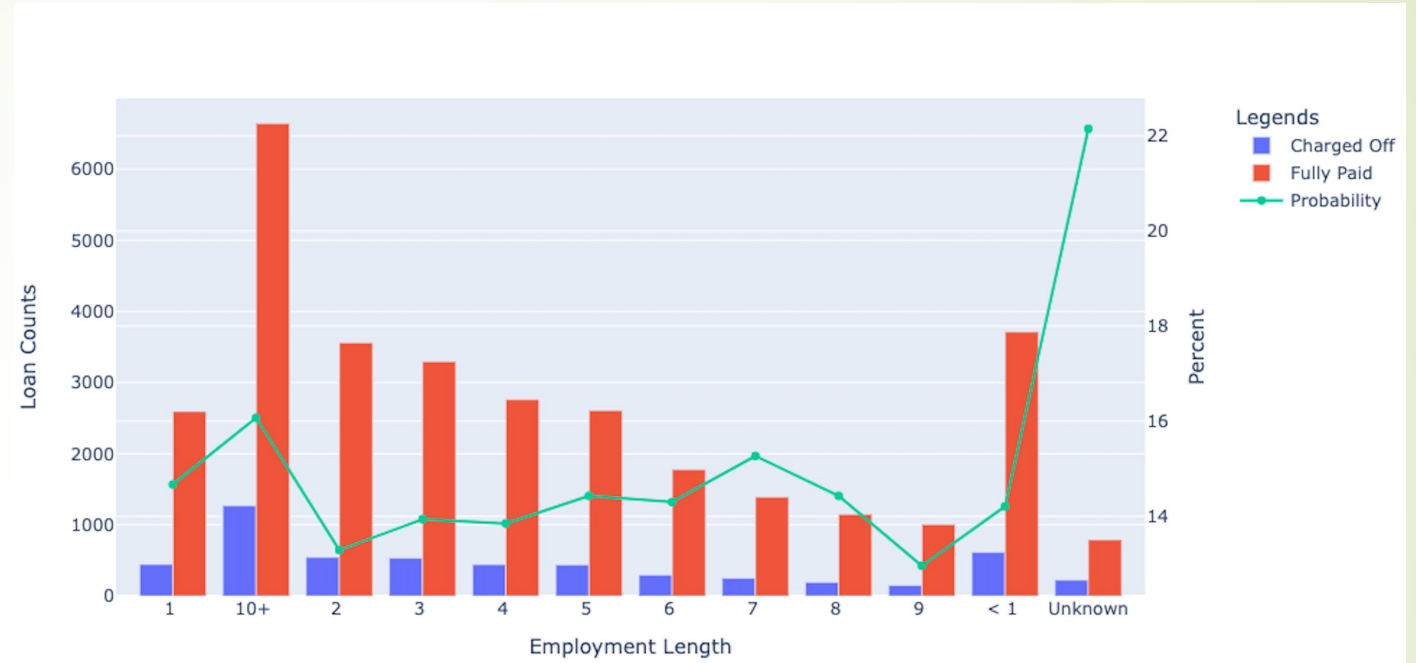
Observation:

➡ Defaulter Probability is high for the applicants living on rent
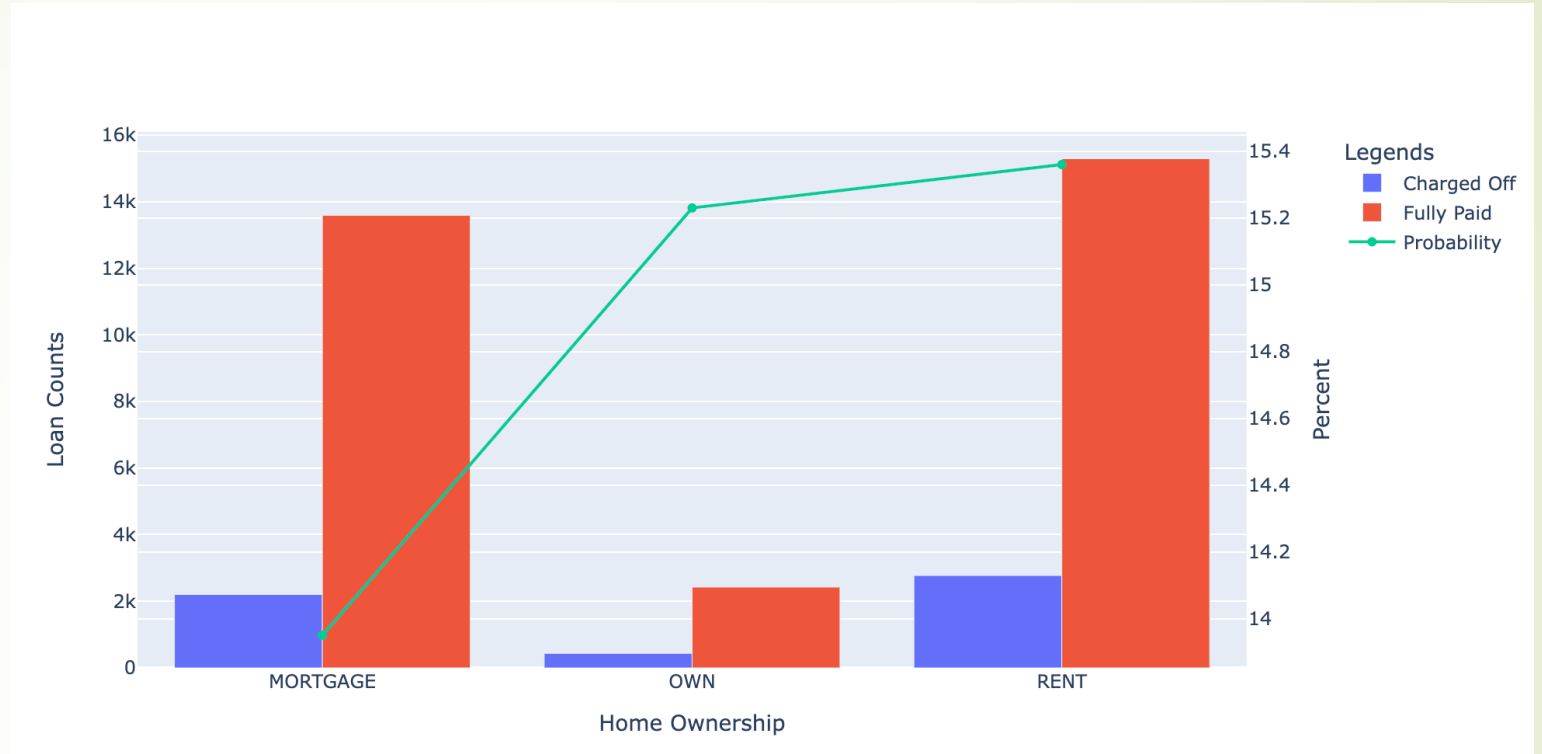
Insights – Probability Analysis

Observation:

➡ Defaulter Probability is high for the applicants having no employment history (Unknown) and with 10+ years of employment length

Insights – Probability Analysis
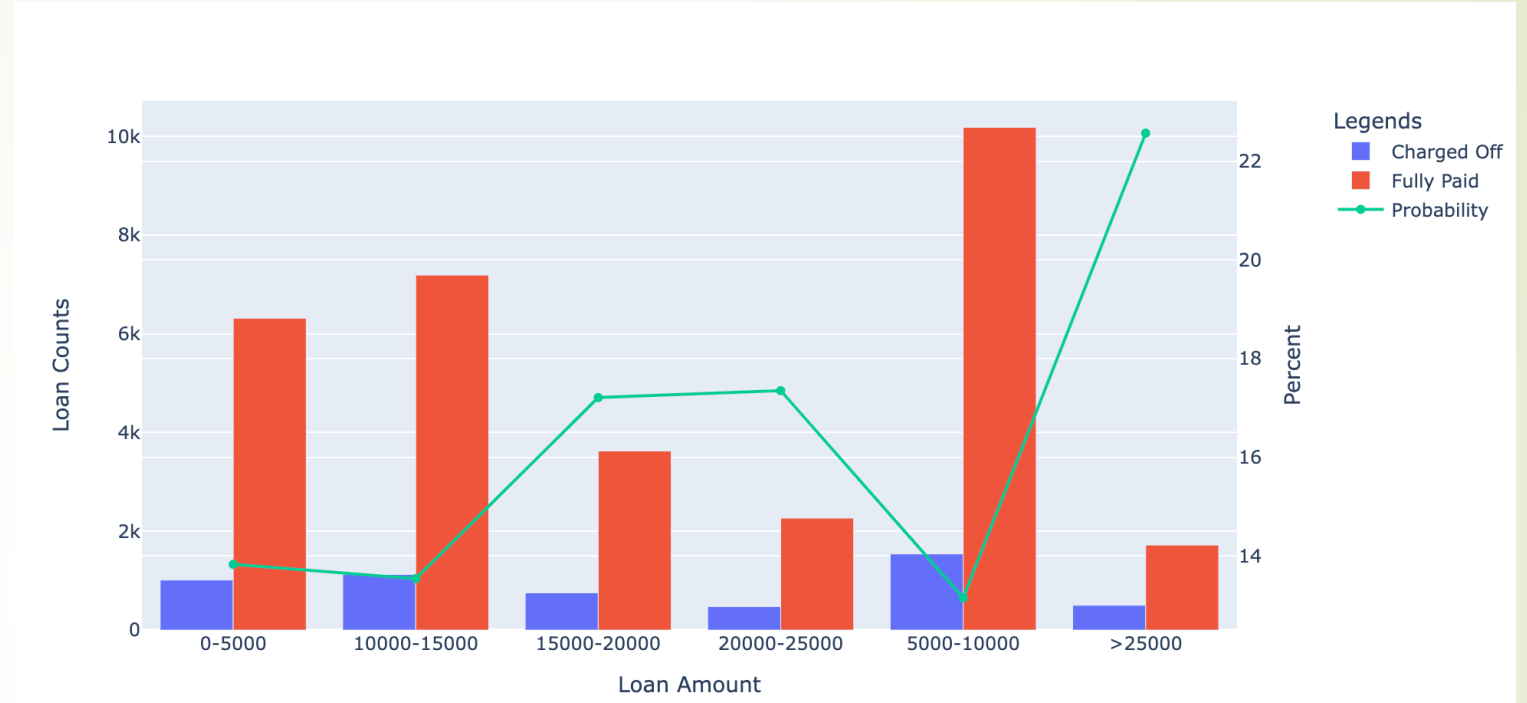
Observation:

- Probability to default is high for applicants running small business
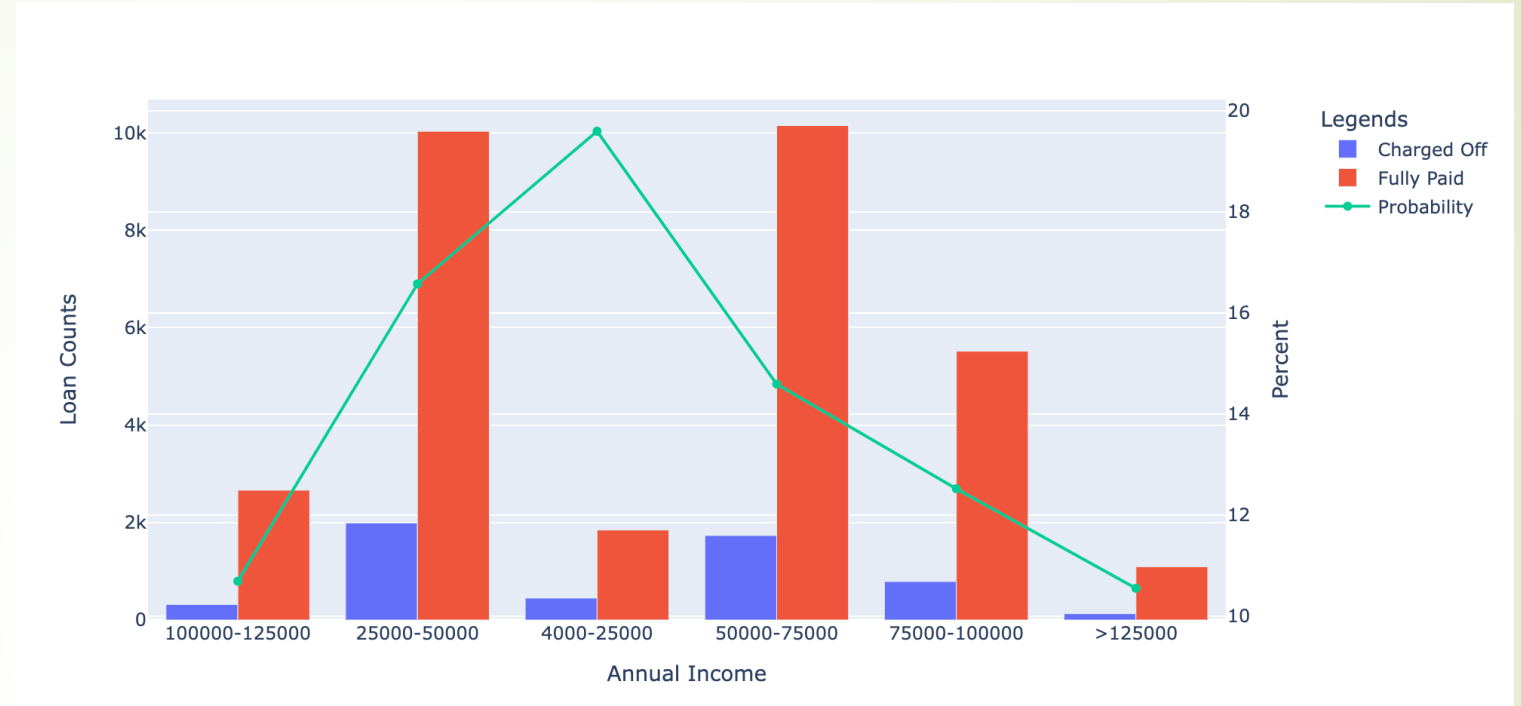
Insights – Probability Analysis

Observation:

- Loan Amount of $25,000 or above has the highest probability to default.

Insights – Probability Analysis

Observation:

➡ Probability of charge off for income range between $4,000-$25,000 is high.

Conclusion

- Applicants who have taken the loan for "small business" are having highest probability (27.79%) to default.

- Defaulter Probability is the highest (22.15%) for the applicants having no employment history (Marked as "Unknown" in the graph)

- Probability to default is increasing from grade A to G. Sub-grade F5 has maximum probability (52.5%) to default.

- Low annual income ($25,000 or less) applicants are highest probable (19.59%) to default.

- Probability to default is increasing with increase in Loan Amount. Applicants with Loan amount of $25,000 or above are highly probable to default.

- Probability to default is increasing with increasing Interest Rate and DTI. Loan applicants with interest rate or DTI 19% or above are highly probable to default.

- Applicants from Nebraska (NE) have the highest probability (60%) to default.

Thank you