

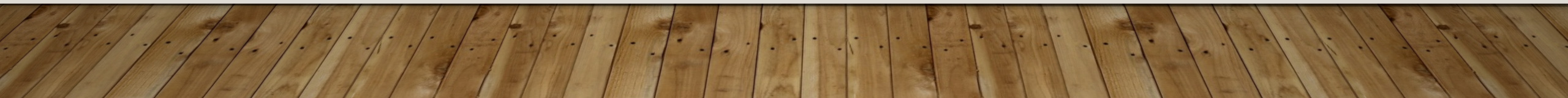
# ANALYSIS AND VISUALIZATION OF LENDING CLUB LOAN DATA

---

DATA CLEANING, OUTLIER DETECTION, AND CORRELATION ANALYSIS

**AUTHORS**

MEGHA KHANT & ANUSHMITA MUKHERJEE



# BUSINESS UNDERSTANDING

---

- Introduction:
  - The company specializes in lending various types of loans to urban customers.
  - Decision to approve loans based on applicant's profile to minimize risk.
- Risks:
  - Loss of business if a good applicant is rejected.
  - Financial loss if a bad applicant is approved.
- Objective:
  - Identify patterns that indicate a person's likelihood to default.
  - Use this information for making informed lending decisions.

# DATA UNDERSTANDING

---

- Dataset Overview:
  - Complete loan data for loans issued from 2007 to 2011.
  - Contains information on loan status: fully paid, current, and charged-off.
- Data Dictionary:
  - Provides descriptions of variables in the dataset.
  - Link to [Data Dictionary](#)

# DATA PREPARATION

---

- Steps Taken:
  - Handling missing values.
  - Removing irrelevant columns.
  - Converting data types for analysis.
- Outcome:
  - Cleaned and prepared dataset ready for analysis.



# UNIVARIATE ANALYSIS

---

- **Objective:** Understand individual variable distributions.
- **Key Findings:**
  - Distribution of loan amounts.
  - Interest rates across loans.
  - Loan purpose distribution.
- **Visualizations:**
  - Histograms and bar plots for key variables.

# BIVARIATE ANALYSIS

---

- **Objective:** Explore relationships between variables.
- **Key Findings:**
  - Relationship between loan amount and interest rates.
  - Default rates by loan purpose.
- **Visualizations:**
  - Scatter plots and box plots showing key relationships.

# MULTIVARIATE ANALYSIS

---

- **Objective:** Understand interactions between multiple variables.
- **Key Findings:**
  - Influence of multiple factors on loan default.
- **Visualizations:**
  - Heatmaps and pair plots.

# KEY INSIGHTS

---

- Driving Factors Behind Loan Default:
  - High loan amounts and interest rates correlate with higher default rates.
  - Certain loan purposes have higher default rates.
- Business Implications:
  - Adjusting interest rates based on risk.
  - Refining loan approval criteria.



# RECOMMENDATIONS

---

- For Loan Approval Process:
  - Implement stricter criteria for high-risk loans.
  - Use higher interest rates for riskier applicants.
- Future Analysis:
  - Consider additional variables like credit scores.
  - Regularly update the analysis with new data.

# CONCLUSION

---

- After doing the EDA on the data set, it indicates that loan defaults are significantly influenced by high loan amounts, high revolving balances, and utilization ratios, along with higher interest rates and instalment payments.
- Borrowers with higher incomes and higher emp\_title able to pay back the loan better, leading to lower default rates.
- Also, home\_ownership and grad are also important factor in loan defaulting.
- Defaulting borrowers are leading to higher recovery costs.