



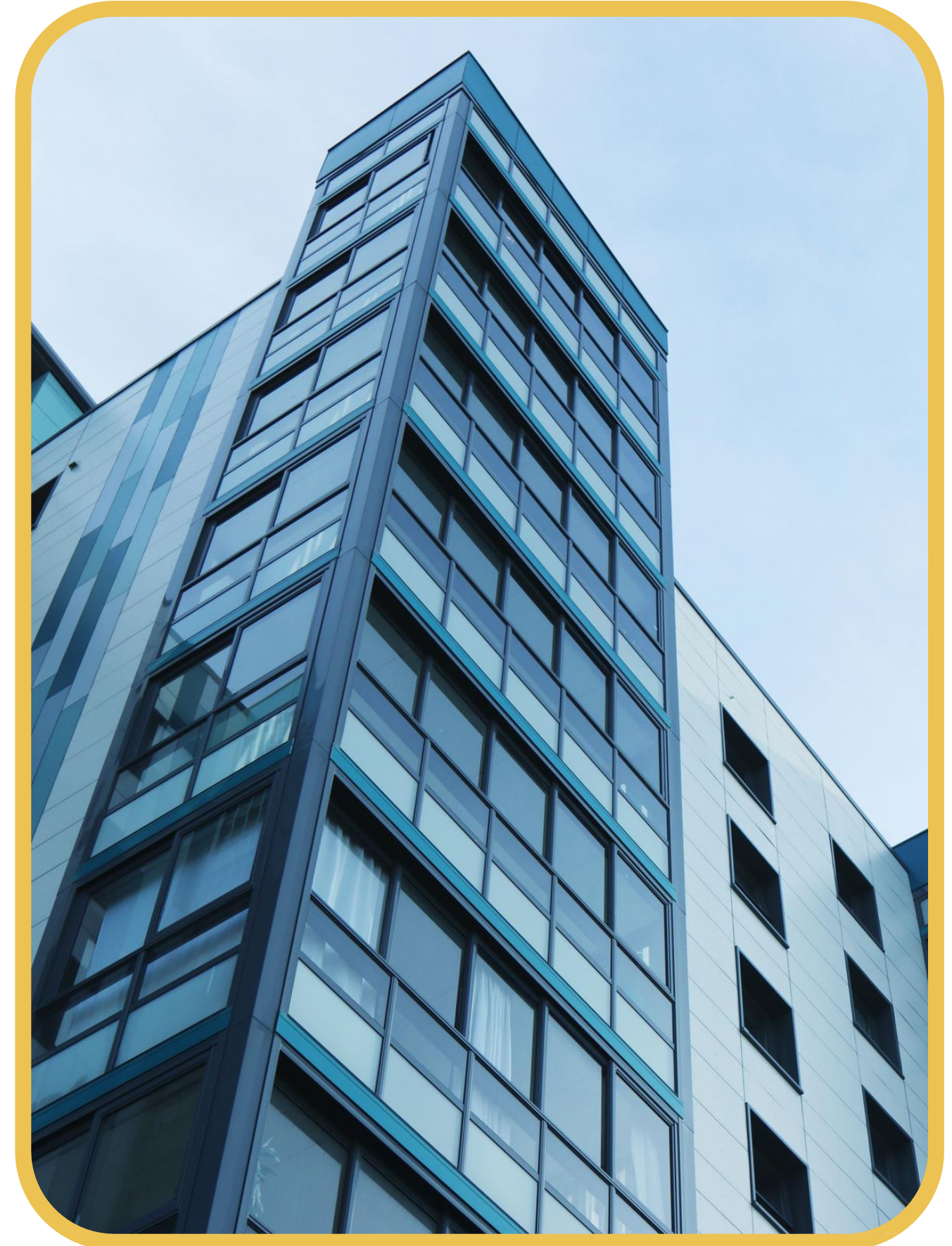
# ANALYSIS AND FINDINGS PRESENTATION

Data Analysis for Loan Default Prediction

# INTRODUCTION

## OBJECTIVE

- To automate the business loan approval process by understanding key factors that lead to loan default.
- Predict the likelihood of a loan default based on historical data.





# DATASET OVERVIEW

- Business loan data containing variables like Loan Amount, Number of Employees, State, City, Loan Status, etc.
- Key columns: MIS\_Status, DisbursementGross, NoEmp, Term, RevLineCr, LowDoc.



# DATA CLEANING AND HANDLING NULL VALUES



## Actions Taken:

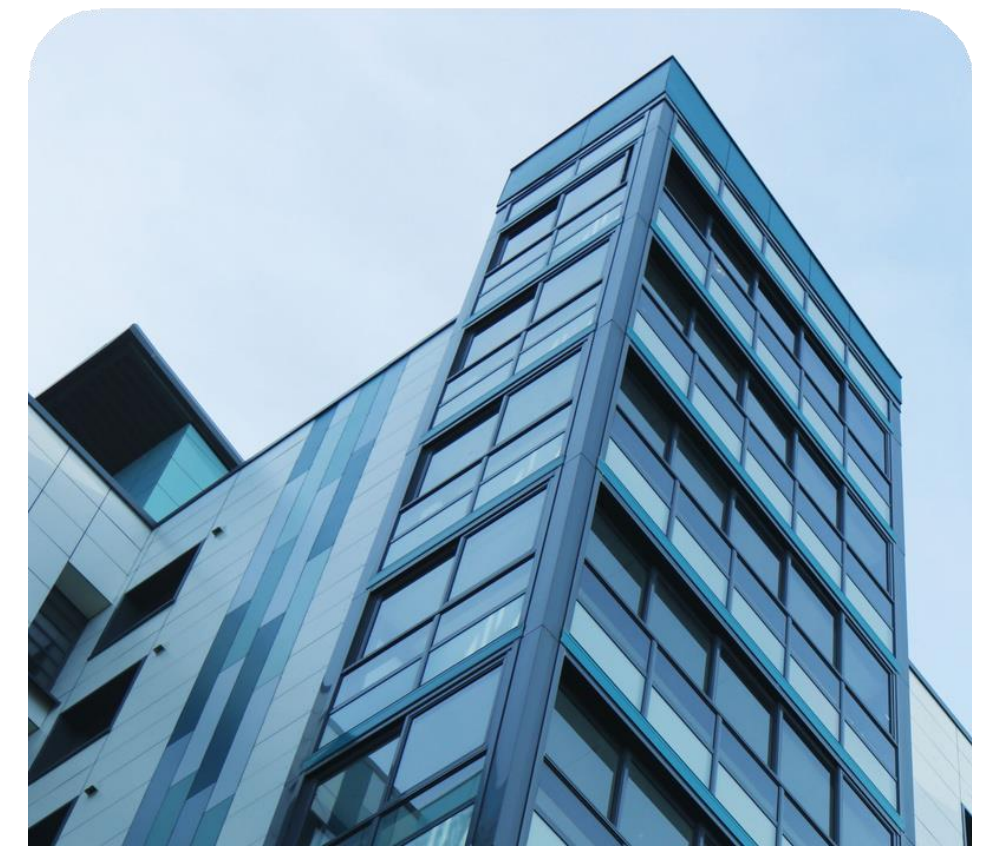
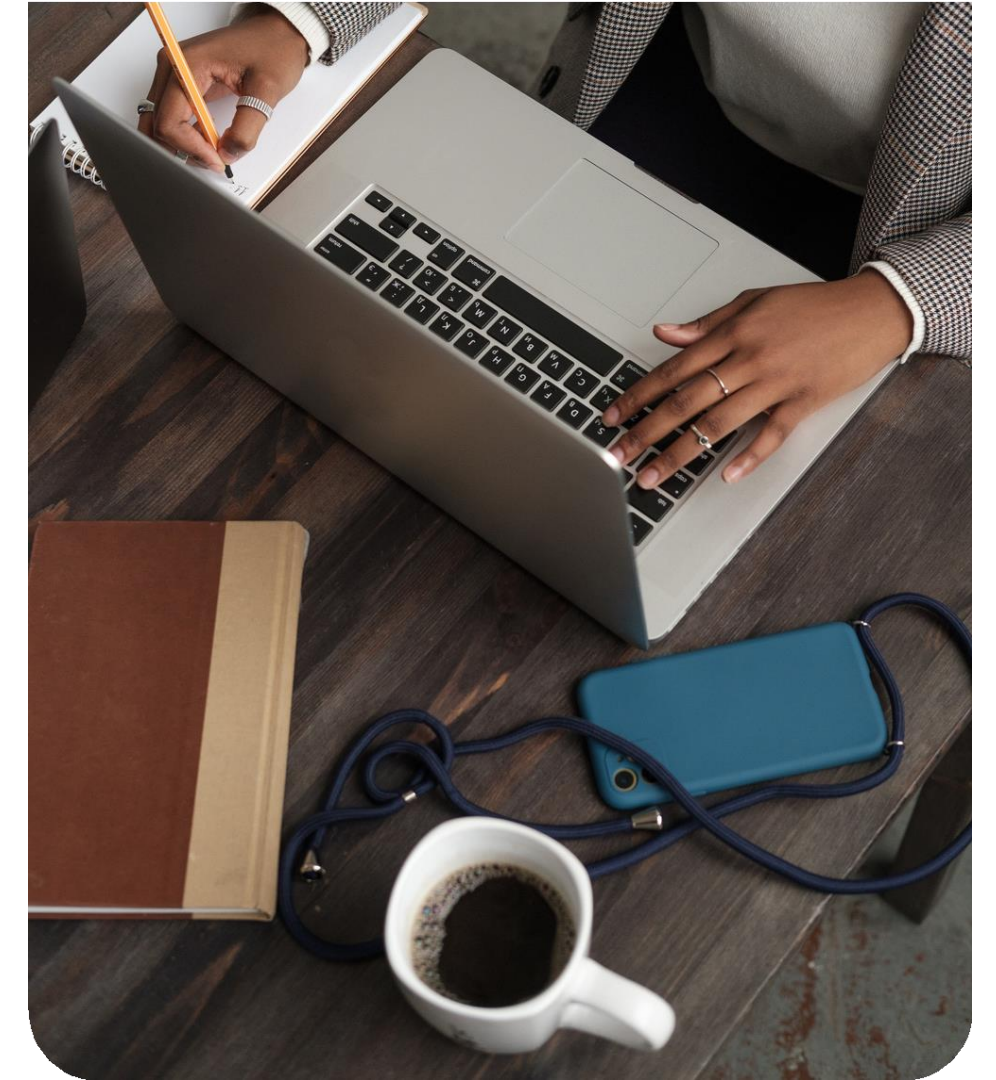
- Handled null values for State, City, Bank, and BankState using ZIP codes.
- Cleaned RevLineCr and LowDoc columns with unexpected values.
- Dropped irrelevant columns (e.g., ChgOffDate).
- Replaced null values in the DisbursementDate column with the mode.
- Deleted remaining null values.



# DATA DISTRIBUTION

## Numerical Column Distributions:

- Loan Amount: Mean = 0.2M, Max = 11M (right-skewed with outliers).
- Remaining Loan: Mean = 2, Max = 0.9M.
- No. of Employees: Mean = 11, Max = 9999.
- Term: Mean = 110 months, Max = 569 months.





# **Exploratory Data Analysis**



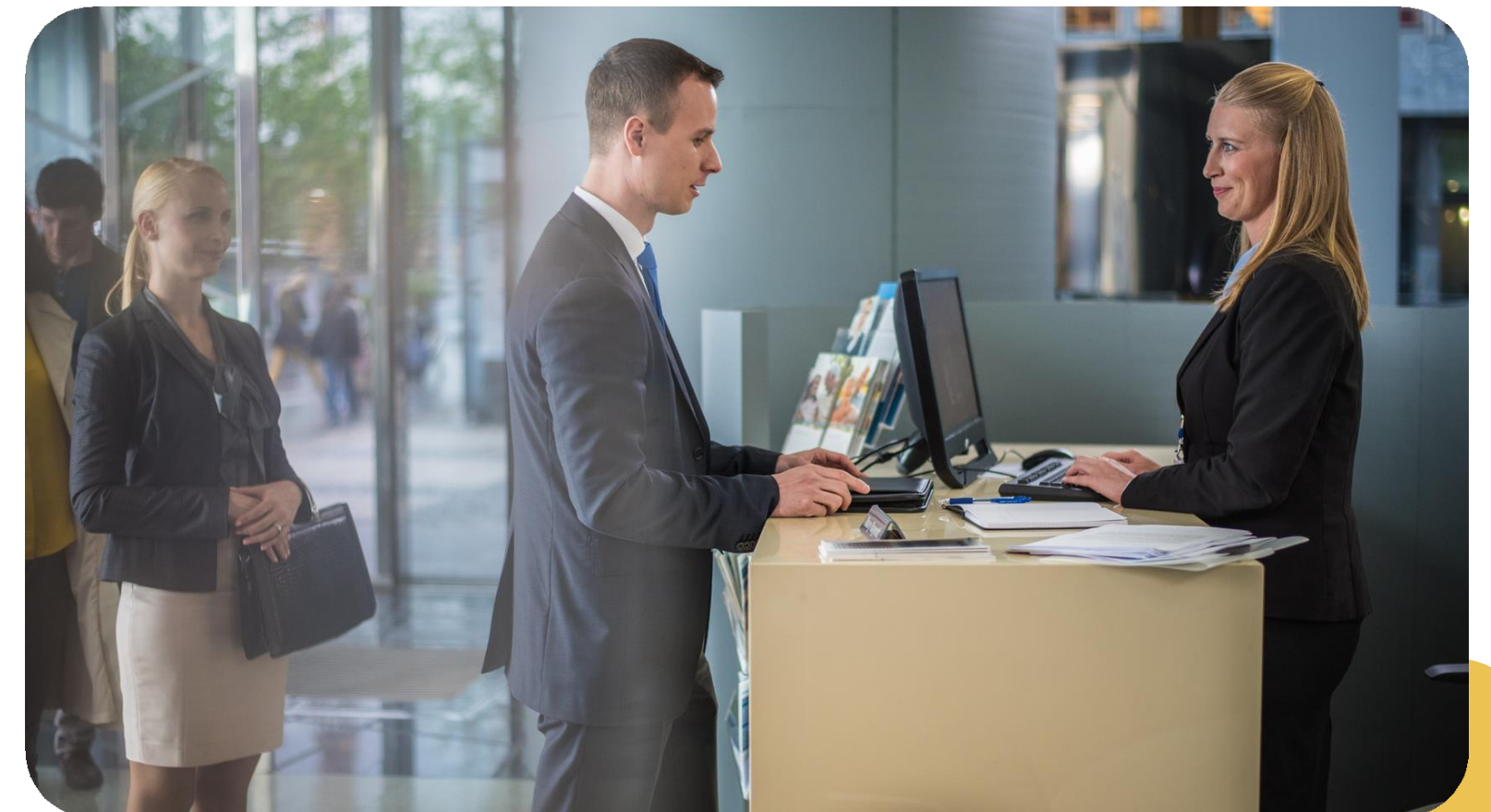


## Top States with Defaulted Loans:

- California: 24,170 defaulted loans.
- City with the most defaults: Los Angeles (3,181).

## Top Bank with Defaulted Loans:

- Bank of America NATL Assoc. 23,948 defaulted loans.



# LOAN AMOUNT AND DEFAULT RISK

## Findings:

- Smaller loans have a higher risk of default.
- Grossly approved loans often correlate with lower risk compared to smaller loans.





# URBAN VS. RURAL LOAN DEFAULTS

## Findings:

- Urban areas: 24% default rate.
- Rural areas: 18% default rate.
- Undefined category: 7% default rate.

# LOW DOCUMENT LOANS

## Findings:

- 9% default rate for low-document loans.
- Consistent with industry trends indicating that low-doc loans carry a higher risk.







# LOAN TERM VS. DEFAULT RATE

## Findings:

- Shorter-term loans ( $\leq 50$  months) have higher default rates.
- Default rates decrease significantly for longer-term loans (100+ months).

# FEATURE ENGINEERING

## New Features Created:

- Term Bucket: Buckets based on loan term.
- Loan Bucket: Buckets based on loan amount.
- Loan Risk Score: A composite score based on loan size, term, and revolving credit status.







# TERM BUCKET ANALYSIS

## Results:

- High-risk bucket: 54% default rate.
- Low-risk bucket: 5% default rate.



# LOAN RISK SCORE

## Risk Score Findings:

- High Loan Risk Score correlates with higher default rates.
- Default rates increase significantly with a higher Loan Risk Score.





# CONCLUSION

## Key Findings:

- Smaller loans and low-document loans show higher risk of default.
- Urban loans tend to default more than rural loans.
- Loan term and revolving credit status are significant predictors of loan performance.





# THANK YOU

By Alifeya Zainuddin



[Alifeya53@gmail.com](mailto:Alifeya53@gmail.com)