INTERNSHIP REPORT
2025-26, SUMMER TERM

Sameer Wanjari
Roll No. 21174028
Engineering Physics

Summer Internship Report

Submitted by:
Sameer Nandesh Wanjari
21174028
IDD Part V, Semester IX
Department of Physics
Indian Institute of Technology (BHU) Varanasi

Table of Contents

About the Company

Introduction

Aham Housing Finance Limited, incorporated on November 21, 2017, is a Non-Banking Financial Company (NBFC) headquartered in Chennai, Tamil Nadu, India, with its registered office in Mumbai, Maharashtra. The company specializes in providing affordable home loans, targeting the informal market and lower economic segments, particularly in metro city outskirts and Tier II/III cities. Aham Housing Finance focuses on enabling homeownership through minimal documentation, attractive interest rates, and streamlined payment processes, supported by technologies like Excel, Python, and Power BI for data-driven operations.

Key Details

- **Industry**: Housing Finance
- **Founded**: 2017
- **Headquarters**: G-122, Phase III, Spencer Plaza, 769, Anna Salai, Chennai, Tamil Nadu - 600002
- **Mission**: To empower customers through accessible home financing and financial literacy.

- **Services**: Home loans (₹10 lakh to ₹15 lakh), minimal documentation options, customer education programs.
- **Technology**: Utilizes Excel, Python, Power BI, ASP.NET, and JavaScript for loan processing and portfolio management.
- **Employee Count**: 130 (as of December 2024)
- **Funding**: Raised $4.87 million from 47 investors, with a valuation of $12.2 million (October 2023).

Aham Housing Finance is committed to addressing complex financial challenges by leveraging data analytics and technology to enhance loan turnaround times, risk management, and customer satisfaction.

Internship Overview

Industrial Training Experience: Analyst Intern at Aham Housing Finance Limited

This report details my 8-week internship as an Analyst Intern at Aham Housing Finance Limited from May 15, 2025, to July 15, 2025. Under the supervision of **Rajkumar P.S.** (Head - Credit), I focused on analyzing loan lifecycle data, automating trackers, and developing dashboards to support data-driven lending and risk management. The internship involved collaboration with Sales, Credit, and Operations teams to enhance loan processing efficiency and generate actionable insights for business and product teams.

Weekly Progress

Week 1: Project Scoping, Data Understanding & Excel Basics

- **Tasks**:
  - **Onboarding & Project Briefing**: Received an overview of Aham Housing Finance's business model, emphasizing affordable home loans and data-driven operations. Understood the project goal: to optimize loan processing and risk analysis.
  - **Introduction to Loan Lifecycle**: Researched key metrics like turnaround time (TAT), login-to-sanction ratios, and delinquency risks in housing finance.
  - **Database Familiarization (Excel)**: Gained access to loan datasets, including customer demographics, loan applications, and repayment records. Explored data structure using Excel.

- ○ **Basic Excel Analysis**: Used Excel functions (e.g., VLOOKUP, COUNTIF) to explore sample data and filter loan statuses.
- **Tools/Things Used & Learnt**:
  - ○ **Software**: Microsoft Excel
  - ○ **Concepts**: Loan lifecycle, housing finance metrics, Excel basics, data exploration.

Week 2: Data Extraction, Cleaning & Initial Exploration with Excel

- **Tasks**:
  - ○ **Data Extraction (Excel)**: Extracted loan data from internal systems into Excel spreadsheets, combining customer details, loan disbursals, and repayment records.
  - ○ **Data Cleaning (Excel)**: Performed initial cleaning tasks, including:
    - ■ Correcting data type inconsistencies (e.g., date formats).
    - ■ Removing duplicates in customer records.
    - ■ Standardizing text fields (e.g., branch names).
  - ○ **Descriptive Statistics (Excel)**: Calculated metrics like average TAT, loan disbursal amounts, and bounce rates using Excel functions (AVERAGE, SUMIFS, COUNTIFS).
  - ○ **Preliminary Visualizations (Excel)**: Created bar and pie charts to visualize loan distribution by branch and customer segment.
- **Tools/Things Used & Learnt**:
  - ○ **Software**: Microsoft Excel
  - ○ **Concepts**: Data extraction, data cleaning, descriptive statistics, basic visualization.

Week 3: Python Setup & Exploratory Data Analysis (EDA)

- **Tasks**:
  - ○ **Python Environment Setup**: Configured a Python environment using Anaconda and Jupyter Notebooks.
  - ○ **Loading Data with Pandas**: Imported cleaned Excel datasets into Pandas DataFrames for advanced analysis.
  - ○ **In-depth Data Cleaning (Python)**: Addressed data quality issues using Pandas:
    - ■ Handled missing values (e.g., imputing average loan amounts).
    - ■ Detected outliers in repayment amounts using IQR methods.
    - ■ Standardized numerical features for analysis.
  - ○ **Comprehensive EDA (Python)**: Conducted univariate and bivariate analysis using Pandas and Matplotlib. Created histograms, box plots, and correlation matrices to identify relationships between loan attributes (e.g., TAT, bounce rates) and customer behavior.
- **Tools/Things Used & Learnt**:
  - ○ **Software**: Python (Jupyter Notebook)

- ○ **Libraries**: Pandas, NumPy, Matplotlib
- ○ **Concepts**: Exploratory Data Analysis, data preprocessing, visualization.

Week 4: Feature Engineering & Loan Metrics Definition

- **Tasks**:
  - ○ **Feature Engineering**: Created new features to enhance analysis, including:
    - ■ **Loan Tenure**: Calculated in months from loan start date.
    - ■ **Average Monthly Repayment**: Derived from repayment logs.
    - ■ **Days Since Last Payment**: Measured recency of customer activity.
    - ■ **Bounce Count**: Number of failed repayment attempts.
  - ○ **Defining Key Metrics**: Established metrics like delinquency risk (e.g., loans with multiple bounces) and login-to-sanction ratio for analysis.
  - ○ **Data Aggregation (Excel)**: Aggregated data for complex metrics (e.g., total bounces per customer) using Excel pivot tables.
  - ○ **Feature Selection**: Identified relevant features for risk analysis based on EDA findings.
- **Tools/Things Used & Learnt**:
  - ○ **Software**: Python (Pandas), Microsoft Excel
  - ○ **Concepts**: Feature engineering, metric definition, data aggregation, domain knowledge.

Week 5: Predictive Modeling - Risk Analysis Introduction

- **Tasks**:
  - ○ **Introduction to Risk Analysis**: Learned about statistical and machine learning approaches for predicting delinquency risks in housing finance.
  - ○ **Data Splitting**: Split loan datasets into training (80%) and testing (20%) sets for model evaluation.
  - ○ **Baseline Model Training**: Trained a basic Logistic Regression model using Scikit-learn to predict delinquency risk based on features like bounce count and repayment history.
  - ○ **Model Evaluation Metrics**: Calculated Accuracy, Precision, Recall, and F1-score to assess model performance.
- **Tools/Things Used & Learnt**:
  - ○ **Software**: Python (Jupyter Notebook)
  - ○ **Libraries**: Scikit-learn
  - ○ **Concepts**: Classification, training/testing split, model evaluation metrics.

Week 6: Model Optimization & Advanced Evaluation

- **Tasks**:
  - ○ **Hyperparameter Tuning**: Adjusted parameters (e.g., regularization strength for Logistic Regression) to improve model performance.

- ○ **Cross-Validation**: Applied k-fold cross-validation to ensure model robustness.
- ○ **Advanced Metrics**: Used ROC-AUC and confusion matrices to evaluate model performance, addressing imbalanced data (e.g., low delinquency rates).
- ○ **Feature Importance**: Identified key predictors of delinquency (e.g., bounce count, days since last payment) using model outputs.
- **Tools/Things Used & Learnt**:
  - ○ **Software**: Python
  - ○ **Libraries**: Scikit-learn
  - ○ **Concepts**: Hyperparameter tuning, cross-validation, ROC-AUC, feature importance.

Week 7: Insights & Dashboard Development

- **Tasks**:
  - ○ **Insights Extraction**: Translated model outputs into business insights, e.g., "Customers with over 2 bounces and no payments in 30 days are 3x more likely to default."
  - ○ **Retention Strategies**: Collaborated with the Credit team to propose strategies like targeted follow-ups for at-risk customers and improved documentation processes.
  - ○ **Dashboard Development (Power BI)**: Developed Power BI dashboards to visualize TAT, login-to-sanction ratios, and delinquency risks by branch and customer segment.
  - ○ **Scenario Analysis (Excel)**: Estimated the impact of proposed strategies on loan recovery rates using Excel scenarios.
- **Tools/Things Used & Learnt**:
  - ○ **Software**: Python, Microsoft Excel, Power BI
  - ○ **Concepts**: Business insights, dashboard design, scenario planning.

Week 8: Report Generation, Presentation & Future Scope

- **Tasks**:
  - ○ **Comprehensive Report Writing**: Compiled this report, detailing methodologies, findings, and recommendations for loan process optimization.
  - ○ **Presentation Development**: Prepared a PowerPoint presentation summarizing key insights and recommendations for non-technical stakeholders.
  - ○ **Future Work**: Suggested enhancements like real-time delinquency monitoring and integrating additional data (e.g., customer credit scores).
  - ○ **Final Review**: Presented findings to Rajkumar P.S. and the Credit team, incorporating feedback into deliverables.
- **Tools/Things Used & Learnt**:
  - ○ **Software**: Microsoft Word, Microsoft PowerPoint
  - ○ **Skills**: Technical writing, presentation skills, strategic recommendations.

Key Skills Acquired and Enhanced

- **Data Analysis**: Advanced Excel functions, Python (Pandas, NumPy) for data manipulation, and statistical analysis.
- **Data Visualization**: Created reports and dashboards using Power BI and Matplotlib for loan metrics and risk insights.
- **Machine Learning**: Supervised learning (classification), model training, evaluation, and hyperparameter tuning with Scikit-learn.
- **Database Management**: Data extraction and analysis using Excel and Python, with conceptual understanding of data warehousing.
- **Problem Solving**: Addressed data quality issues and proposed data-driven solutions for loan processing and risk management.
- **Communication**: Presented technical findings to diverse audiences and wrote comprehensive reports.
- **Project Management**: Managed multiple tasks, prioritizing deliverables within tight timelines.