Swift Loan Data Review & Action Plan

This presentation outlines our approach to ensure data quality in our loan platform, leading to accurate reporting, informed decision-making, and improved efficiency. We will review current data quality, define key metrics, and propose an action plan.

by Werner Koegelenberg

Index

1. Introduction

Objective: Evaluate Swift Loan’s suitability for $1 million in debt capital investment.

Role of Data Analysis:

- Validate loan tape data quality.
- Derive actionable insights and recommendations.
- Support the investment team with key performance metrics.

Swift Loan's Opportunity: Showcase RealFi’s analytical capabilities to support their business performance.

2. Process

Collection and Preparation

The first step in data analysis involves gathering raw data. Once collected, the data must be cleaned to address missing values, outliers, and inconsistencies, ensuring its accuracy and reliability. This step also includes organising and formatting the data, such as normalizing values and converting types, to make it suitable for analysis. Proper preparation ensures the foundation for a successful and meaningful analysis.

Exploration and Analysis

In this step, the focus shifts to exploring the prepared data through descriptive statistics and visualizations to identify patterns, trends, and correlations. Analytical techniques such as regression, clustering, or hypothesis testing are applied to uncover deeper insights and relationships within the data. This process helps analysts transform raw data into meaningful information, setting the stage for drawing conclusions and making data-driven decisions.

Interpretation and Communication

The final step is to interpret the analytical findings and communicate them effectively. This involves summarizing key insights and presenting them clearly through visualizations like charts, graphs, or dashboards. Analysts also provide actionable recommendations based on their results, ensuring stakeholders can understand and use the information to inform strategies and decision-making. This step bridges the gap between data insights and practical applications.

3. Rationale

Collection and Preparation

- Download the supplied files from Google Cloud as supplied.
- Create a file structure that would support storage for the review, analysis, coding scripts, and explanations (readme files) for the duration of the project.
- Create Private **GitHub** repository and share it with the relevant team members to ensure collaboration.

Exploration and Analysis

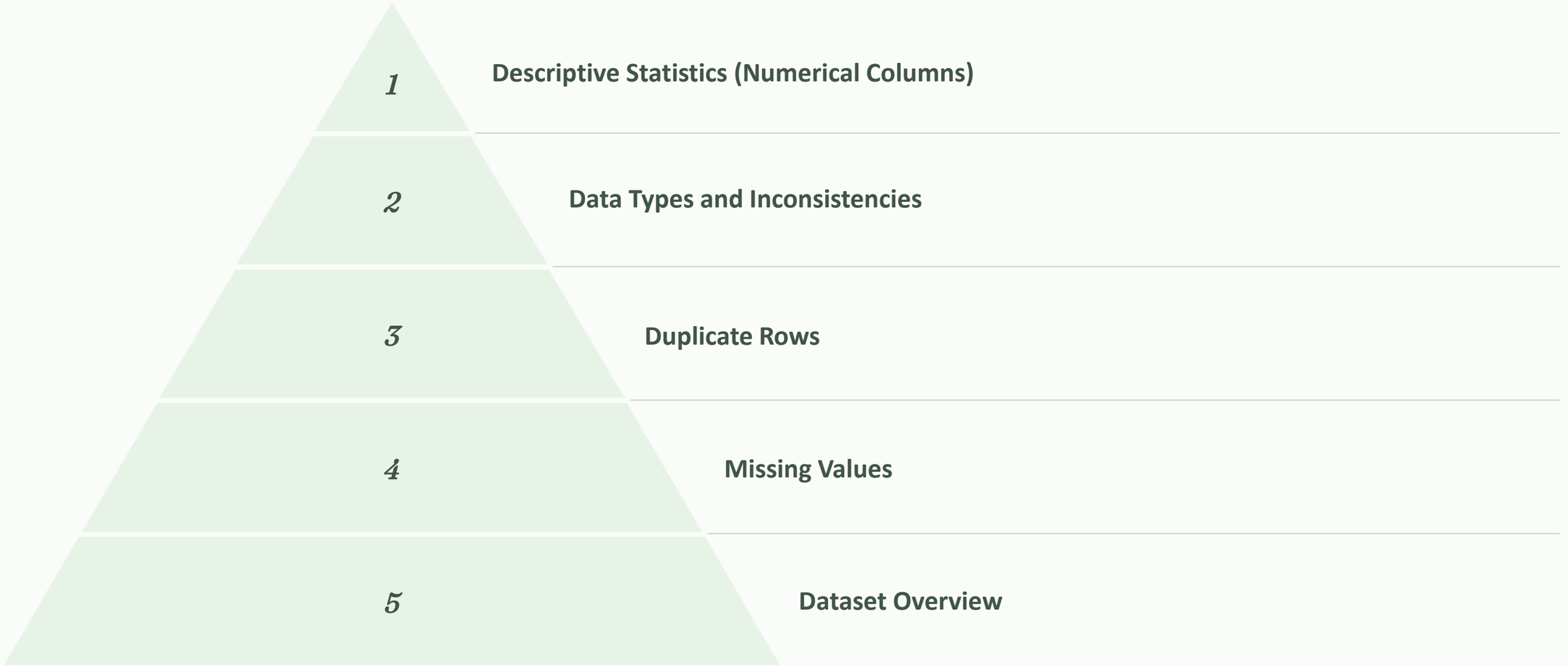
- Using **Python** in **VSCode** write a simple script that would extract 20 rows of each dataset for general investigation.
- Explore the sample data extractions with MS Excel
- Write a reusable Python script to generate data quality reports per data source.
- Set up a cloud environment in **AWS** to support storage with **AWS S3**
- Use **AWS Glue** to create data catalog, validate data quality and clean data
- Create **ETL** (Extract, Transform, Load) pipelines in AWS Glue to Clean and Join data
- Use **AWS Athena** to query, check, validate data by means of **SQL** queries

Interpretation and Communication

- **AWS QuickSight** was used to develop visual representations of the data
- Compile **MS PowerPoint** presentation to share findings
- Communication by Email

4. Data Quality Findings

Our script (data_quality_analysis.py) analyzes the loan tape data and generates an initial report per data table shared. The focus of the analysis is broken down below.



4. Data Quality Findings continued..

In the initial review of the data, we identified some key findings.

Borrowers Dataset (borrowers.csv):

- A large number of missing ‘geo_classification’ values (not critical for analysis)
- Incorrectly classified data types (**critical for analysis**)

- **Loans Dataset** (loans.csv):
- A number of missing ‘borrower_id’ values. (**statistically makes up 1.8% of unique values**)

- **Payments Dataset** (payments.csv):
- Some incorrectly classified data types (**critical for analysis**)
- Two missing ‘loan_id’ values. (statistically makes up 0.00015% of unique values)

5. Plan of Action

To ensure the successful management of data quality going forward we would like to suggest:

1 Data Completeness

Ensure that checks are in place for user fields, ensuring that the user can not complete an application before all fields are completed.

2 Data Accuracy

Implement a data management system that will automatically generate key system fields through automation.

3 Data Consistency

To ensure uniform data formats, develop ETL pipelines that aligning with the established data dictionary.

5. Plan of Action continued…

Short Term

1.Implement Data Cleaning Procedures

Priority: High

Action: Remove duplicates and inconsistencies in key fields by means of automation scripts.

Correct data types and handle missing values by means of automation scripts.

2.Standardize Data Formats

Priority: High

Action: Standardize currency, date, and numeric formats.

3.Establish Data Validation Rules

Priority: Medium

Objective: Prevent future data errors.

Action: Implement validation rules for data entry.

Medium Term

1. **Implement Data Quality Monitoring and Alerts:**

   Priority: Medium

   Reason: Enable proactive identification of data quality issues as they arise.

   Action: Set up automated monitoring tools with alerts for key data quality metrics, such as missing values and outliers.

2. **Develop a Data Governance Framework:**

   Priority: Medium

   Reason: Establish clear roles, responsibilities, and processes for maintaining data quality.

   Action: Create a data governance framework that includes policies for data ownership, stewardship, and quality assurance.

3. **Invest in Data Quality Training and Tools:**

   Priority: Low

   Reason: Equip staff with the skills and tools necessary to maintain high data quality standards.

   Action: Provide training on data quality best practices and invest in data management tools that facilitate data cleaning and validation.

6. Benefits of Improved Data Quality

Accurate Reporting

Generate reliable and insightful reports for decision-making.

Enhanced Risk Management

Identify potential risks and vulnerabilities more effectively.

Improved Customer Experience

Offer better service by providing accurate and timely information.

7. Portfolio Performance Metrics

To clean and join data we used AWS Glue. Scripts provided in GitHub repo. Examples of tables pre- and post cleaning.

Tables (5)

Last updated (UTC)
November 23, 2024 at 10:50:55

Delete

View and manage all available tables.

Filter tables

Name	Database	Table fo...	Location	Classific...	Last Updated (UTC)
payments	raw_data	-	s3://realfi-loan-tape-data-nov-2024/rawData/payments/	CSV	November 23, 2024 at 06:20:16
loans	raw_data	-	s3://realfi-loan-tape-data-nov-2024/rawData/loans/	CSV	November 23, 2024 at 06:19:42
borrowers	raw_data	-	s3://realfi-loan-tape-data-nov-2024/rawData/borrowers/	CSV	November 23, 2024 at 06:17:30
combined_loans_payments	processed_data	-	s3://realfi-loan-tape-data-nov-2024/processedData/	Parquet	November 23, 2024 at 06:47:12
combined_borrowers_loans	processed_data	-	s3://realfi-loan-tape-data-nov-2024/processedData/	Parquet	November 23, 2024 at 06:34:57

7a. Portfolio Performance Metrics

Write off rate

The write-off rate was calculated to assess the portion of the total loan principal that could not be collected and was subsequently written off as a loss. To determine this rate, we first filtered the dataset to include only loans with a maturity date on or before 31 July 2024, ensuring that the analysis focused on loans that had reached the end of their term.

Currency	Write-Off Rate (%)
KES	36,45%
UGX	30,21%

Write off rate per Gender

Currency	Gender	Collection Rate (%)
KES	Female	38,87%
KES	Male	32,15%
UGX	Female	38,38%
UGX	Male	27,05%

7b. Portfolio Performance Metrics

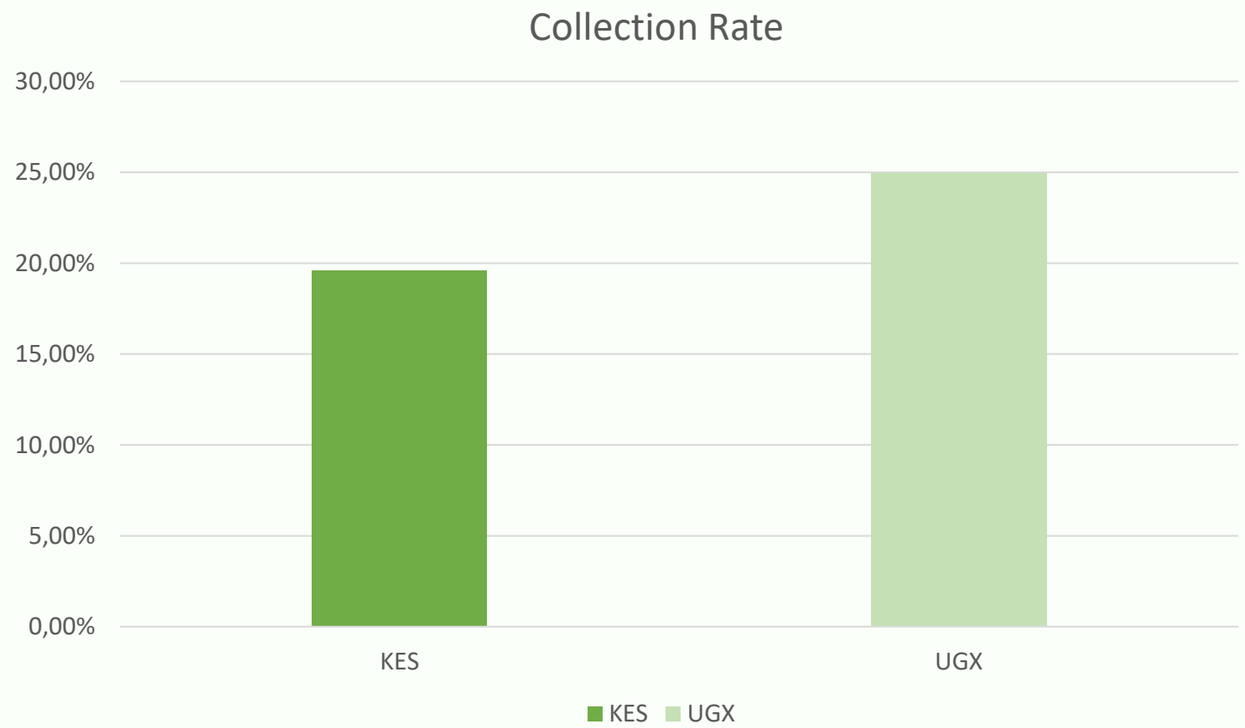
Collections rate

The Collections rate was calculated as Total Write-Offs / Total Loan Volume for both genders, rural and urban filtered by country.

Currency	Collections Rate (%)
KES	19,56%
UGX	24,93%

Based on the results loans issued as UGX has a better collection rate than KES.

****Note**, *last collected payment across currencies were noted to be 31 July 2024, therefore we considered only calculating the collection rate per currency for loans maturing before 31 July 2024.*



7b. Portfolio Performance Metrics continued..

Collection rate – per Gender

Research shows that woman score better on loan repayments in Africa.

Ref:

- https://www.businessdailyafrica.com/bd/economy/women-score-better-than-men-in-loan-repayment-4715160
- https://www.un.org/africarenewal/magazine/august-2015/loans-women-smart-economics

Hypothesis: Focus on female lenders to increase collection rates.

Calculation: Total Collection Amount divided by the Total Principal Amount by currency and gender for loans that matured before 31 July 2024.

Currency	Gender	Collection Rate (%)
KES	Female	20,54%
KES	Male	19,37%
UGX	Female	23,13%
UGX	Male	23,38%

Insights

Gender Disparity: In both currencies (KES and UGX), male borrowers have lower collections rate compared to female borrowers. This could indicate differences in repayment behavior or financial capabilities between genders.

Collections Rate Variance: The collections rates vary significantly between currencies. Understanding the underlying reasons for these differences could help in tailoring financial products or strategies to improve collections.

Potential Areas of Focus: Swift Loan might want to explore targeted interventions to assist borrowers Kenya in improving their repayment performance, given the lower collections rates observed.

7c. Portfolio Performance Metrics

Gross Yield

In order to calculate the gross yield we had to first determine the loan period. Interest income is calculated by multiplying the principal amount by the interest rate, adjusted for this period. Payments collected are aggregated to understand the total amount received for each loan.

Principal, interest, penalties, fees, and payments—are combined to compute total income for each loan. Gross yield is then calculated by dividing total income by the principal amount, resulting in a percentage that indicates profitability.

Country	Gross Yield(%)
Kenya (KES)	7,37%
Uganda (UGX)	3,67%

Insights

In Interest Rates: The disparity might be due to higher interest rates on KES loans, resulting in greater income relative to the principal.

Loan Terms: Differences in loan terms (such as duration and repayment schedules) could contribute to the higher yields in KES.

Economic Conditions: Economic conditions in Kenya might be more favorable for borrowers, leading to better loan performance and, consequently, higher yields.

Borrower Profiles: The profile of borrowers in each currency could also differ, with KES borrowers potentially representing sectors or demographics with higher repayment capabilities.

7d. Portfolio Performance Metrics

Average Days in Arrears

The table provides insights into various time intervals associated with loans, particularly focusing on default and write-off processes.

Description	Number of Days
Average Loan Period (days):	28,72
Average Days From Issue date to Default date:	90,00
Average Days From Issue to Write-off (for loans with write-offs):	101,41

Insights

- On average, loans default around 90 days after they are issued. This suggests that, while the official loan term is about 28.72 days, borrowers may face challenges meeting their obligations, leading to defaults occurring significantly later. This could indicate issues with borrower financial stability or the terms being misaligned with borrower capacities.
- The fact that write-offs occur, on average, 11.41 days after defaults (101.41 days from issue minus 90 days from issue to default) suggests that the organization moves relatively quickly to write off loans once they default, possibly indicating efficient collection and write-off procedures.

8. Insights

Based on the analysis conducted on the loan data for Kenyan Shillings (KES) and Ugandan Shillings (UGX), several key insights and recommendations emerge:

High Write-off Rates: The write-off rates are notably high for both KES and UGX loans, with KES loans experiencing an 36,45% rate and UGX loans a 30,21% rate. This indicates a significant challenge in loan recovery, where a large proportion of the principal amount has been deemed uncollectible.

Gross Yield Disparity: Loans in KES exhibit a higher average gross yield (7,37%) compared to UGX (3,67%). Despite the yields, the profitability is overshadowed by the elevated write-off rates.

Early Defaults: The average number of days between default and maturity suggests that defaults are occurring well before loans reach their maturity, signaling potential issues in borrower creditworthiness or economic pressures affecting repayment capabilities.

Short Loan Periods: The average loan period is relatively short, around 29 days, which may limit borrowers' ability to repay comfortably and contribute to the high default and write-off rates.

9. Recommendations

Enhance Risk Assessment: Strengthening risk assessment and borrower screening processes could help identify potential default risks early on, allowing for more informed lending decisions and tailored loan terms.

Improve Collection Strategies: Implementing more robust collection strategies, including proactive borrower communication and support, could mitigate write-offs. Exploring flexible repayment plans or restructuring options might enhance recovery rates.

Review Loan Terms: Consider extending loan durations slightly to provide borrowers with more time to manage repayments, potentially reducing default rates and improving loan performance.

Focus on Economic Conditions: Conduct an in-depth analysis of the macroeconomic factors impacting borrowers in both currencies. Understanding these conditions could guide strategic adjustments in lending practices and risk management.

Granular Data Analysis: Collect and analyze more granular data to differentiate between rural and urban clients, male and female borrowers, and various age segments. This segmentation can provide insights into demographic influences on loan performance and help tailor products and services to specific groups.

Borrower Education Programs: Develop education programs to better inform borrowers about the terms and expectations of their loans. Clearer understanding and communication can empower borrowers to manage their finances more effectively and meet repayment obligations.

Diversify Loan Portfolio: To mitigate risks, financial institutions might benefit from diversifying their loan portfolio across different sectors or borrower profiles, reducing dependency on high-risk segments.

By addressing these areas, financial institutions can work towards reducing write-offs, enhancing loan performance, and improving overall financial health. These insights and recommendations provide a roadmap for strategic improvements in loan management and risk mitigation.

10. Conclusion

In conclusion, the analysis of the loans data by Swift Loan in Kenyan Shillings (KES) and Ugandan Shillings (UGX) reveals significant challenges in loan recovery, highlighted by high write-off rates and early defaults. Despite the higher gross yields observed for KES loans, the profitability is undermined by the elevated write-off rates, suggesting the need for strengthened risk assessment, improved collection strategies, and potentially extended loan terms to enhance repayment capabilities.

To address these challenges, financial institutions should focus on more granular data analysis to understand demographic influences on loan performance, such as differences between rural and urban clients, gender, and age segments. Additionally, implementing borrower education programs to clarify loan terms and expectations can empower borrowers to manage their finances effectively, thereby reducing default rates. By incorporating these strategies, institutions can work towards improving loan performance and overall financial health.

Thank You

GitHub Repo: https://github.com/Lucrotech/data_analyst_takehome_test
(feel free to request access)
AWS Resources: can be shared on request