




Loan Data Analysis

Insights on Customer Base & Prepayments (Sondertilgung)
Patterns

Presentation by

Solomon Uvere Okoro

<https://github.com/soulsuv/data-cleaning-and-analysis-with-SQL-in-Snowflake>



Introduction



Case study to analyze the Customer data from a loan business model.

Data overview:


“fmBKDetail_combined” (BK_Detail): 3619 rows of data on the origination of the loan: credit_scores: A-H, with A as the best score.

“fmKonto_combined” (Konto): 3619 rows of data with information on the current balances of the accounts (as at 15.04.2024).

“fmBuchung_combined” (Buchung): 84020 rows of data with information on each booking in the system (coupled with a dictionary for details on the booking types).

“fmKunde_combined” (Kunde): 3617 rows of data with information on customers (anonymised).


Data Entry: Data entered in snowflake with proper data types definition for all columns.



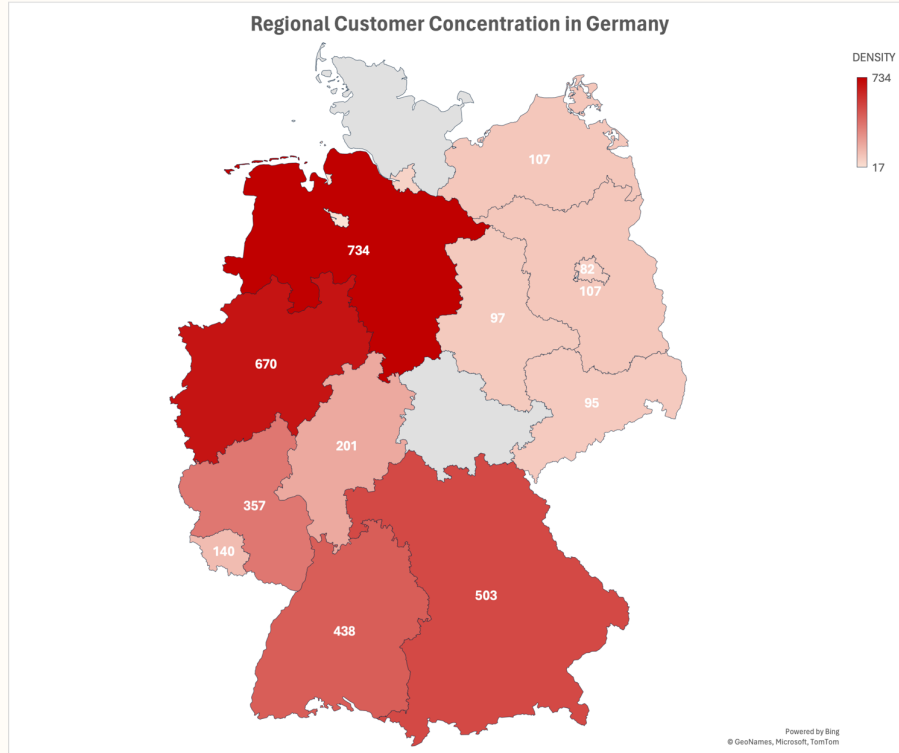
Data Cleaning



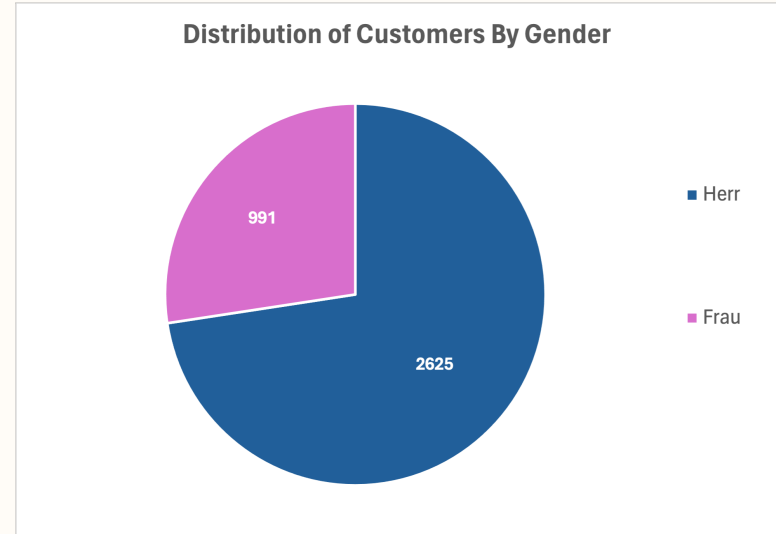
General approach: Check for consistency and anomalies, remove errors and invalid outliers, perform imputation where possible for missing data, perform validation and update Tables.

1. BK_Detail Table: Deleted one row of data with invalid Account_Number and inconsistent entries.
 2. KONTO TABLE: a) Deleted one row of data with invalid Account_Number and inconsistent entries. b) Reset the SPV_TAG data to "En_bv" and "GF_Alpha".
 3. Kunde Table: a) Deleted one row of data with invalid Account_Number/Customer_Number and inconsistent entries. b) Imputed Postal_Codes for rows with Town names as Postal_Codes. c) Imputed country code as "DE" given Town names were in Germany.
 4. Buchung Table: a) Deleted 58 rows of data with invalid Account_Number. b) Concatenated the two parts of Account Number Columns to form the key: Account_Number. c) Reset the Booking_Status encoding from 1-18 to Text. d) Observed and removed 272 duplicate rows in the Buchung Table (total of 330 rows removed from this Table.)
 5. Joined the data using the Account_Number column from all Tables.
- 

Customer Base (Kunde) Demographics Analysis

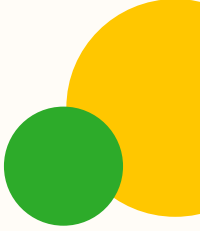


Drivers: Population, economic activity, sunlight exposure, infrastructure and availability, and development rate.

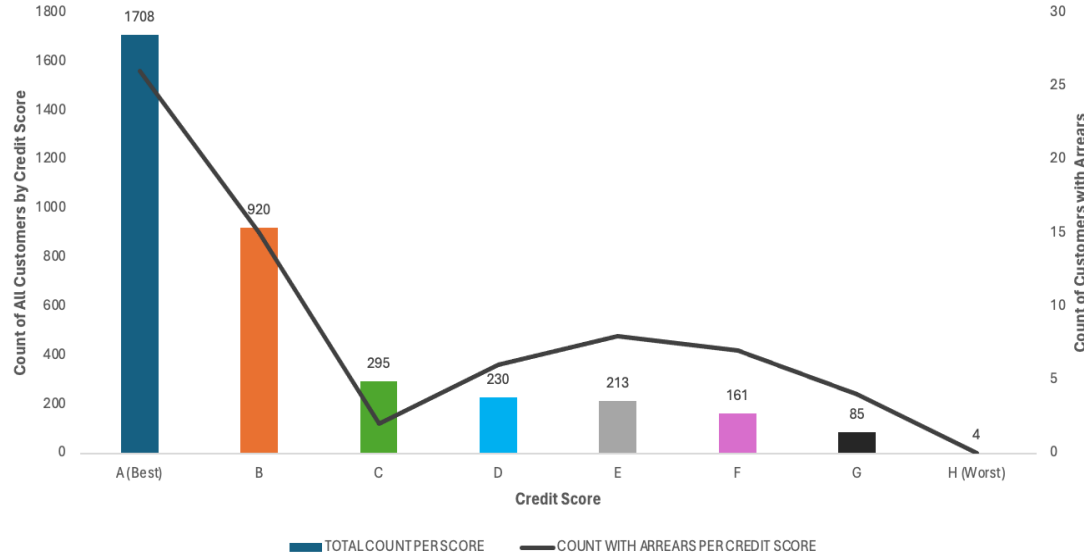


Dominated by Males despite slightly more female in Germany. Further analysis on customers as family units.

Customer Base (Kunde) Demographics Analysis (cont'd)

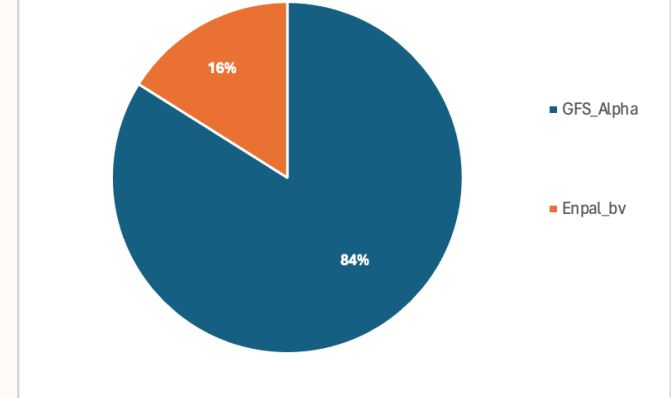


Customer Distribution by Credit Score (Total and For Defaulters)



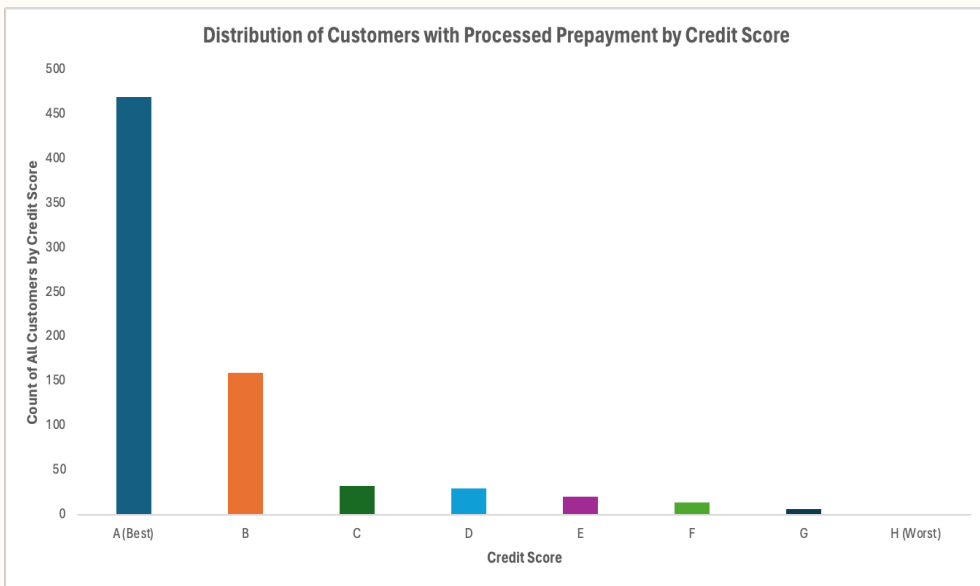
Higher counts of customers with better credit scores shows good screening process for loan applicants. Number of defaulters per Credit Score generally follows the Credit Score distribution. 25% of the 3616 customers had a 0 current principal balance.

Customer Distribution by SPV

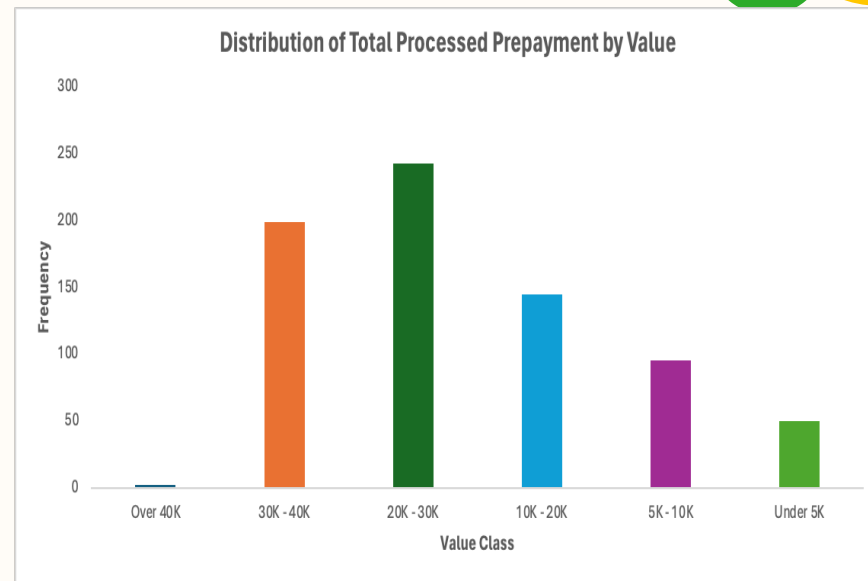


Most of the loans are under the GF_Alpha SPV (84%).

Prepayment Analysis – Sondertilgung Auf Kapital & “Processed Status”



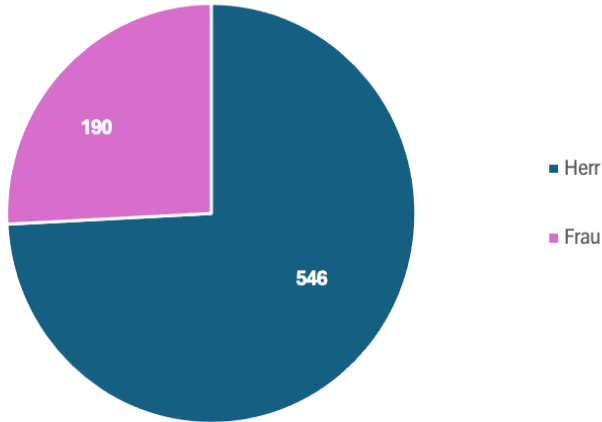
The distribution by credit score for the processed Sondertilgungs Auf Kapital follows the same distribution of the general customer base. (736 Customers).



For Sondertilgungs Auf Kapital that were processed, most transaction values lie between 20K to 40K range.

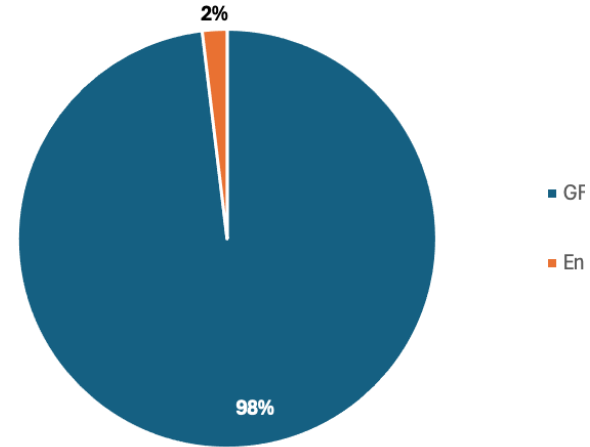
Prepayment Analysis – Sondertilgung Auf Kapital & “Processed Status”

Count of Processed Prepayment by Gender



The distribution by Gender for the processed Sondertilgungs Auf Kapital follows the same distribution of the general customer base.

Count of Processed Prepayment by SPV

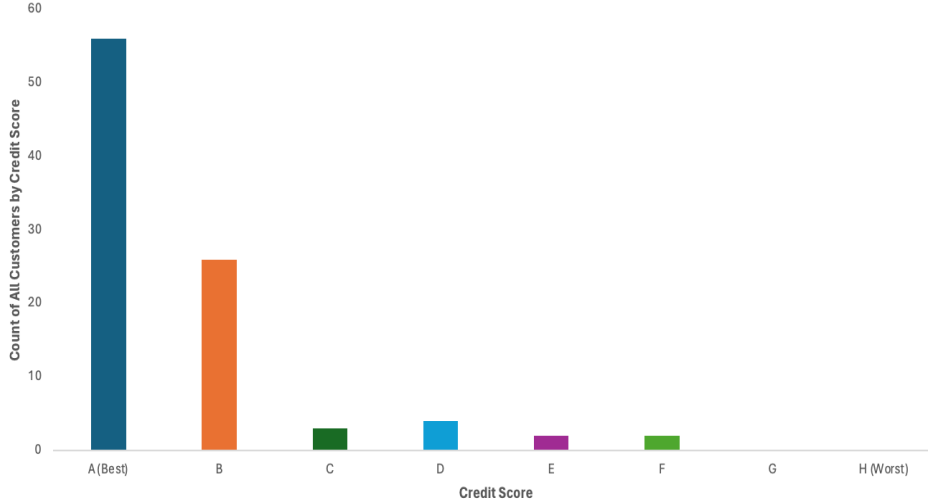


Almost all processed Sondertilgungs Auf Kapital fall under the GF_Alpha SPV.

Prepayment Analysis – Sondertilgung Auf Kapital & “Cancelled Status”

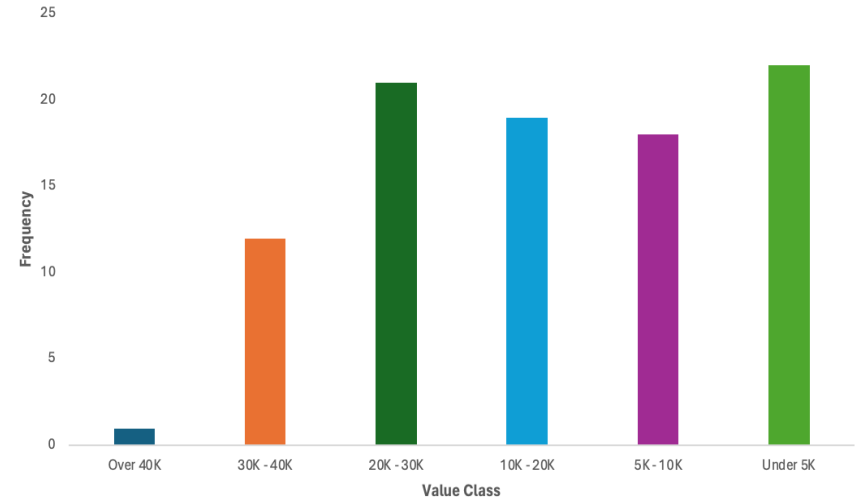


Distribution of Customers with Prepayment and Cancelled Status by Credit Score



The distribution by credit score for the Sondertilgungs Auf Kapital with Cancelled Status follows the same distribution of the general customer base. (93 Customers).

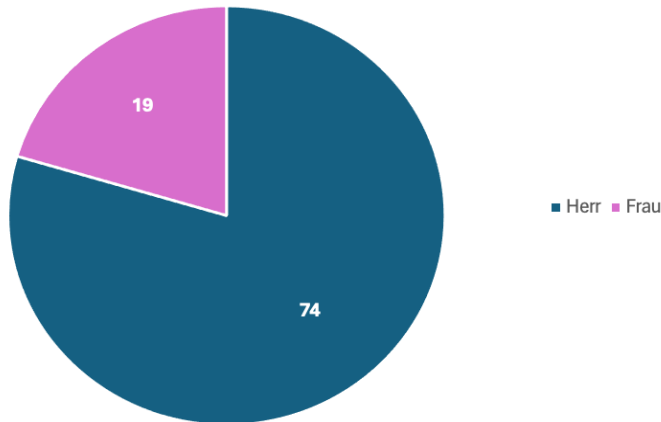
Distribution of Total Cancelled Prepayment by Value



For Sondertilgungs Auf Kapital with Cancelled Status, most transaction values lie between 0K to 40K range with the peak range under 5K.

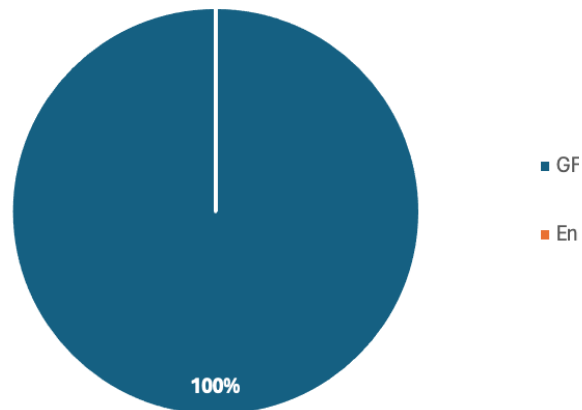
Prepayment Analysis – Sondertilgung Auf Kapital & “Cancelled Status”

Count of Prepayments with Cancelled Status by Gender



The distribution by Gender for the Sondertilgung Auf Kapital with Cancelled Status is slightly skewed in favor of Herrs.

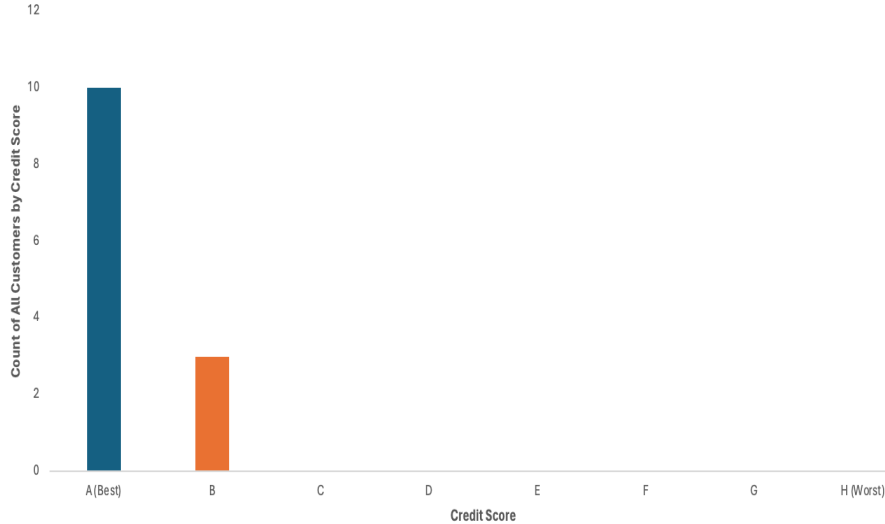
Count of Prepayments with Cancelled Status by SPV



All Sondertilgung Auf Kapital with Cancelled Status fall under the GF_Alpha SPV.

Prepayment Analysis – Storno Sondertilgung

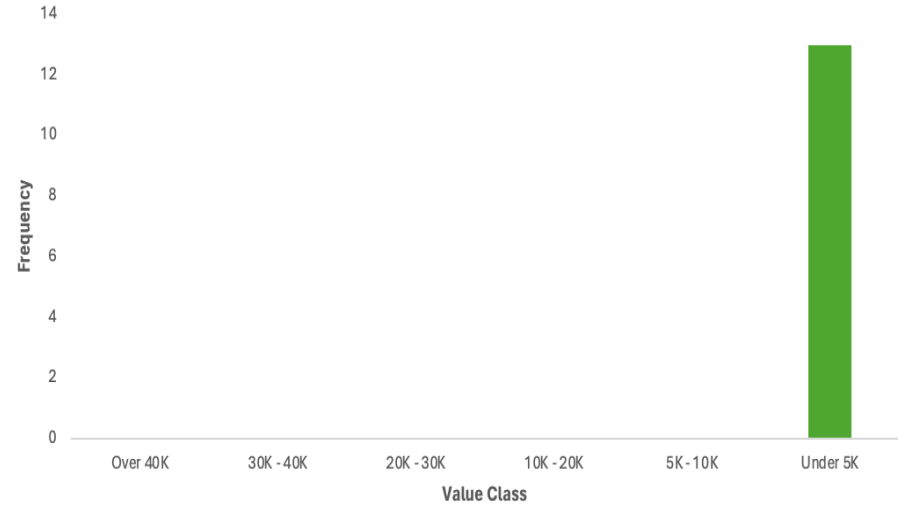
Distribution of Customers with Cancelled Prepayment by Credit Score



Customers with Storno Sondertilgung are 13 in number, with 10 having credit score A and the other 3 with credit score B.

Gender and SPV show similar trends with the general customer data.

Distribution of Total Cancelled Prepayment by Value

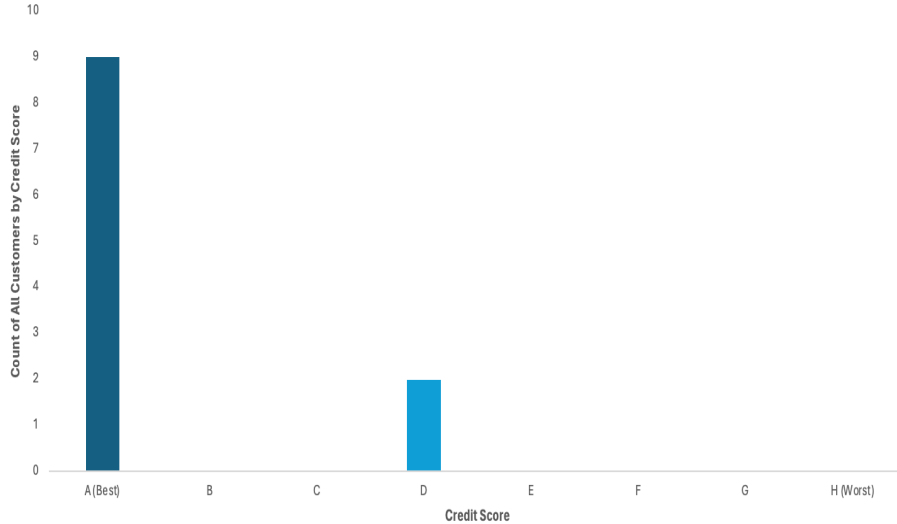


All Storno Sondertilgungs had their transaction values under 5k, (under 200). This is indicative of cancellations stemming from interest payments or other low-value prepayments.

Prepayment Analysis – Storno Sondertilgung Ohne Rückstand

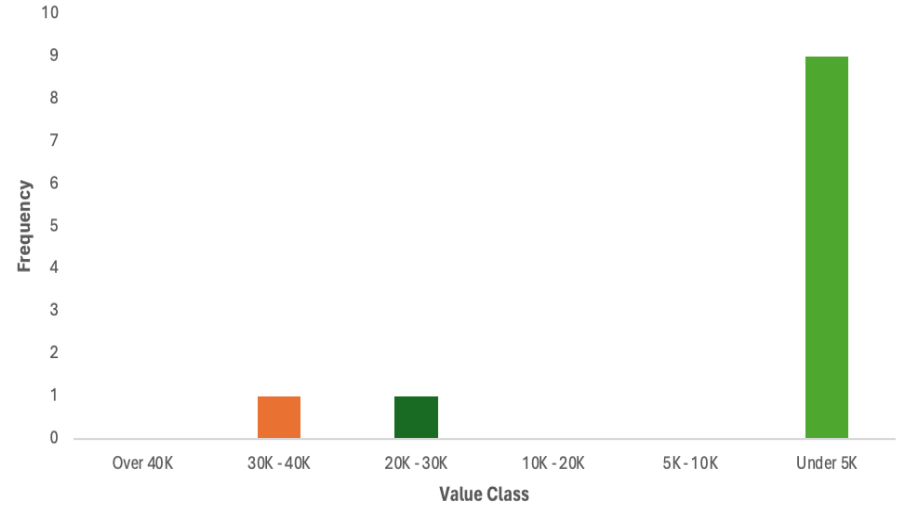


Distribution of Customers with Cancelled Prepayment Without Arrears by Credit Score



11 Customers with Storno Sondertilgung Ohne Rückstand with 9 having credit score A and the other 2 with credit score D.

Distribution of Total Cancelled Prepayment Without Arrears by Value

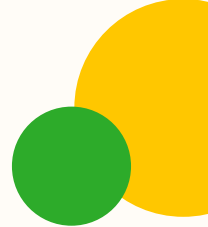


Transaction values for Storno Sondertilgungs Ohne Rückstand are mostly under 5K with a couple values within 20 to 40K.

Gender and SPV show similar trends with the general customer data.


Possible Areas for Further Analysis

- Evaluate the presence of any seasonality.
- Analyse other transaction types present in the Booking.
- Evaluate any regional trends in the analysed data.
- Analyse the time required to settle each transaction to show possible areas of improvement.
- Analyse the nature of the cancellations (Credit or Debits).



Conclusions



- **Effective Data Management:** Comprehensive data cleaning and integration have allowed for accurate analytics.
 - **Creditworthiness and Demographics:** Findings show that the screening process for loan applicants is effective in reducing the number of defaults.
 - **Special Purpose Vehicle (SPV) Utilization:** The majority of loans are managed through the GF_Alpha SPV, which is key to organizing and assessing financial activities efficiently.
 - **Prepayment Behaviour:** Customers typically make larger prepayments, generally between €20K and €40K, suggesting financial stability. The Prepayments with Booking Status as Cancelled need to be further evaluated.
 - **Challenges in Prepayment Cancellations:** Most cancelled prepayments are under €5K, often due to minor issues. This is an area where improvements can be made to enhance customer experience.
 - **Storno Sondertilgung Insights:** The additional analysis of Storno Sondertilgung Ohne Rückstand shows similar demographic trends, with a few higher-value transactions noted. This indicates a need for targeted customer communication and support to ensure clarity in transaction processing.
- 



Questions?