

Project Journal

Project Title: Data-Driven Insights for Banking: Financial and Consumer Behaviour Analysis

Team Member Name: Suryakarthik

Week 1

Date: 15/12/2024

Task Description:

- Analysed the marketing dataset to understand customer characteristics and trends.
- Performed data manipulation for columns such as housing, loan, and default, converting categorical strings into Boolean formats.

Time Spent: 5 hours

Challenges or Difficulties Encountered:

- Ensuring consistency in Boolean conversions while maintaining data interpretability.

Solutions or Actions Taken:

- Utilized lambda functions and Pandas .map() for consistent transformations.
- Cross-validated results by analysing column distributions (value counts()) before and after transformation.

Learning Outcomes:

- Gained hands-on experience in data cleaning and preprocessing using Python.
 - Developed a deeper understanding of categorical data transformation techniques.
-

Week 2

Date: 26/12/2024

Task Description:

- Corroborated metadata for the default dataset to ensure consistency and reliability.
- Verified unique values for columns like SEX, EDUCATION, and MARRIAGE against documentation to resolve discrepancies.
- Created visualizations to explore patterns in the marketing and default datasets.

Time Spent: 7 hours

Challenges or Difficulties Encountered:

- Unclear documentation for certain categorical values in the EDUCATION column.
- High variability in payment status (PAY_0, PAY_2, etc.) columns.

Solutions or Actions Taken:

- Mapped undocumented categories (EDUCATION values 5 and 6) to custom labels (e.g., "Unknown-5").
- Filtered and summarized payment statuses to identify frequent patterns and unknowns.

Learning Outcomes:

- Improved ability to interpret and address undocumented or ambiguous data categories.
- Enhanced skills in using Pandas and Seaborn for exploratory visualizations.

Week 3

Date: 04/01/2025

Task Description:

- Developed detailed heatmaps for the default dataset to visualize relationships between age, credit limit, and default status.
- Interpreted trends in heatmaps to identify key patterns:
 - High credit limit concentrations around specific "steps" (e.g., TWD 50,000, 200,000).
 - Clear age ranges associated with higher default frequencies.
 - Worked on writing the **Conclusion and Future References** section of the report to summarize key findings and suggest potential areas for further research.

Time Spent: 7 hours

Challenges or Difficulties Encountered:

- Heatmap binning required careful calibration to ensure readability.
- High-frequency credit limit bins (e.g., TWD 50,000) obscured minor trends.

Solutions or Actions Taken:

- Adjusted bin sizes for both age and credit axes to balance detail and readability.
- Annotated observations in heatmaps for better team communication.

Learning Outcomes:

- Improved proficiency in creating and interpreting 2D heatmaps with Matplotlib and Seaborn.
 - Learned techniques to manage visual complexity and highlight significant insights.
-

Week 3

Date: 05/01/2025

Task Description:

- Edited and Merged the video presentation of all team members, ensuring a cohesive and professional final product.
- Contributed to scripting, transitions, and visuals for the team's video presentation.

Time Spent: 4 hours

Challenges or Difficulties Encountered:

- Aligning the presentation styles and content flow of different team members.

Solutions or Actions Taken:

- Used video editing software to create seamless transitions and added annotations to emphasize key points.
- Regularly consulted team members to ensure the video met everyone's expectations.

Learning Outcomes:

- Improved video editing and team collaboration skills.
 - Gained a deeper appreciation for multimedia communication in presenting complex projects.
-

Weekly Summary

Total Time Spent: 22+ hours

Key Accomplishments:

- Cleaned and transformed categorical data in the marketing and default datasets.
- Corroborated metadata for the default dataset and resolved undocumented categories.
- Designed heatmaps to identify patterns in age and credit distribution for defaulted and non-defaulted clients.
- Authored the **Conclusion and Future References** section, summarizing insights and proposing areas for further research.
- Merged and edited the team's video presentation, ensuring it was cohesive and engaging.