**Project Coversheet**

| Full Name | Valarmathi Ganessin |
| --- | --- |
| Email | valarsri@gmail.com |
| Contact Number | 0044 7903122243 |
| Project Title (Example – Week1, Week2, Week3) | Week 4 Advanced AI vs Python |

**Project Guidelines and Rules**

**1. Formatting and Submission**

* Format: Use a readable font (e.g., Arial/Times New Roman), size 12, 1.5 line spacing.
* Title: Include Week and Title (Example - Week 1: Travel Ease Case Study.)
* File Format: Submit as PDF or Word file
* Page Limit: 4–5 pages, including the title and references.

**2. Answer Requirements**

* Word Count: Each answer should be within 100–150 words; Maximum 800–1,200 words.
* Clarity: Write concise, structured answers with key points.
* Tone: Use formal, professional language.

**3. Content Rules**

* Answer all questions thoroughly, referencing case study concepts.
* Use examples where possible (e.g., risk assessment techniques).
* Break complex answers into bullet points or lists.

**4. Plagiarism Policy**

* Submit original work; no copy-pasting.
* Cite external material in a consistent format (e.g., APA, MLA).

**5. Evaluation Criteria**

* Understanding: Clear grasp of business analysis principles.
* Application: Effective use of concepts like cost-benefit analysis and Agile/Waterfall.
* Clarity: Logical, well-structured responses.
* Creativity: Innovative problem-solving and examples.
* Completeness: Answer all questions within the word limit.

**6. Deadlines and Late Submissions**

* Deadline: Submit on time; trainees who fail to submit the project will miss the “Certificate of Excellence”

**7. Additional Resources**

* Refer to lecture notes and recommended readings.
* Contact the instructor or peers for clarifications before the deadline.

| **YOU CAN START YOUR PROJECT FROM HERE** |
| --- |

**Week 4 Advanced AI Vs Python**

### Project Objective Comparison: No-Code vs. Coding

**Clean** – Handle missing values, remove duplicates

**Preprocess** – Prepare data with transformations and categorization

**Visualize** – Create charts and summary visuals

**Predict** – Build or run predictive models

**Evaluate** – Measure model accuracy and performance

**Compare** – Analyze results across multiple models

**Automate** – Repeat tasks using templates or loops

**Extract** – Derive insights and trends

**Interpret** – Understand model logic and variable impact

**Strategize** – Use insights for planning and forecasting

**Communicate** – Share results through visuals and summaries

**Storytell** – Present data-backed narratives visually and clearly

### Google AutoML cannot be used with 500 rows because it requires at least 1,000 for analysis, whereas Power BI works perfectly with no minimum data limit.

Task 1: AI-Powered Data Cleaning and Preprocessing PowerBI Vs Python

| **Power BI(Codeless)** | **Python(Code)** |
| --- | --- |
| 1.Upload the Dataset:  **Data Quality Overview**   * **Validity**: % of entries that follow the expected format (e.g., 100% valid for Age, Gender) * **Errors**: % of entries that are incorrectly formatted (e.g., 1% error in Date) * **Missing Values**: % of empty/null entries (e.g., 10% missing in Income) | Statistics for: Purchase\_Frequency  Mean: 15.35  Median: 16.0  Standard Deviation: 8.47532667952388  Min: 1  Max: 29  Statistics for: Seasonality  Skipped — not numeric  Statistics for: Sales  Mean: 54378.954  Median: 54032.5  Standard Deviation: 27263.10646776021  Min: 5203  Max: 99835 |
| 2: **Handling Missing Values**  Replaced column null values with median  Date → 01/01/2024  Income →85375.5  Credit\_Score →588.5  Loan\_Amount→29817 | Replaced date → mode  Replaced Income → Median  Replaced Credit\_Score→ Mean  Replaced Loan\_Amount → Mode  Script used  df['Date'] = df['Date'].fillna(df['Date'].mode()[0])  df['Income'] = df['Income'].fillna(df['Income'].median())  df['Credit\_Score'] = df['Credit\_Score'].fillna(df['Credit\_Score'].mean())  df['Loan\_Amount'] = df['Loan\_Amount'].fillna(df['Loan\_Amount'].mode()[0]) |
| 3: **Detect and Handle Outliers**    No outliers found. | ✅ No outliers found in column: \*\*Customer\_ID\*\*  ✅ No outliers found in column: \*\*Age\*\*  ✅ No outliers found in column: \*\*Income\*\*  ✅ No outliers found in column: \*\*Spending\_Score\*\*  ✅ No outliers found in column: \*\*Credit\_Score\*\*  ✅ No outliers found in column: \*\*Loan\_Amount\*\*  ✅ No outliers found in column: \*\*Previous\_Defaults\*\*  ✅ No outliers found in column: \*\*Marketing\_Spend\*\*  ✅ No outliers found in column: \*\*Purchase\_Frequency\*\*  ✅ No outliers found in column: \*\*Sales\*\*  ✅ No outliers found in column: \*\*Customer\_Churn\*\*  📌 Outliers detected in column: \*\*Defaulted\*\*  Defaulted  6 1  9 1  11 1  12 1  16 1  .. ...  486 1  490 1  491 1  492 1  497 1  [95 rows x 1 columns] |
| 4: **Save the Cleaned Data** |  |

**Task 2: AI-Powered Data Visualisation and Storytelling**

| **Sales Performance over time** from **January to April 2024**. It highlights:   * **Fluctuations in total sales** across dates * **Peaks** like 0.91M in early January and 0.57M in March * **Troughs** like 0.04M in February and 0.08M in April | These anomalies are due to the aggregation of Dates. Sales peak in January. |
| --- | --- |
| Shows February 1, 2024 has medium seasonality shopping. | **Thursday, February 1, 2024**, with the *Medium* group clearly dominating. |
| is highlighting a surprising pattern in data for **Monday, January 15, 2024**:   * It compares **average Spending Score** across three **Seasonality** levels: Low, Medium, and High. * **Low Seasonality** customers had the **highest average Spending Score** on that day. * **High Seasonality** customers—who usually spend more during peak times—actually spent the least. | Low Seasonality Group having the Highest Spending Score on 15/01/2024 |
| The number of customers is **fairly evenly distributed** across all three Seasonality groups  There’s a **slight peak** in the *High Seasonality* group, but nothing extreme | There is relatively even distribution of customers by seasonality |
| Customers in the **Low Seasonality** group have the **highest average Credit Score**  Those in the **Medium Seasonality** group have the **lowest** | Relatively Even Credit\_Score by Seasonality |
| Customers in the **Low Seasonality** group are, on average, **older**   * + Those in the **Medium Seasonality** group are the **youngest** | Relatively Even Age by Seasonality |

Step 3: Use AI Features for Deeper Insights

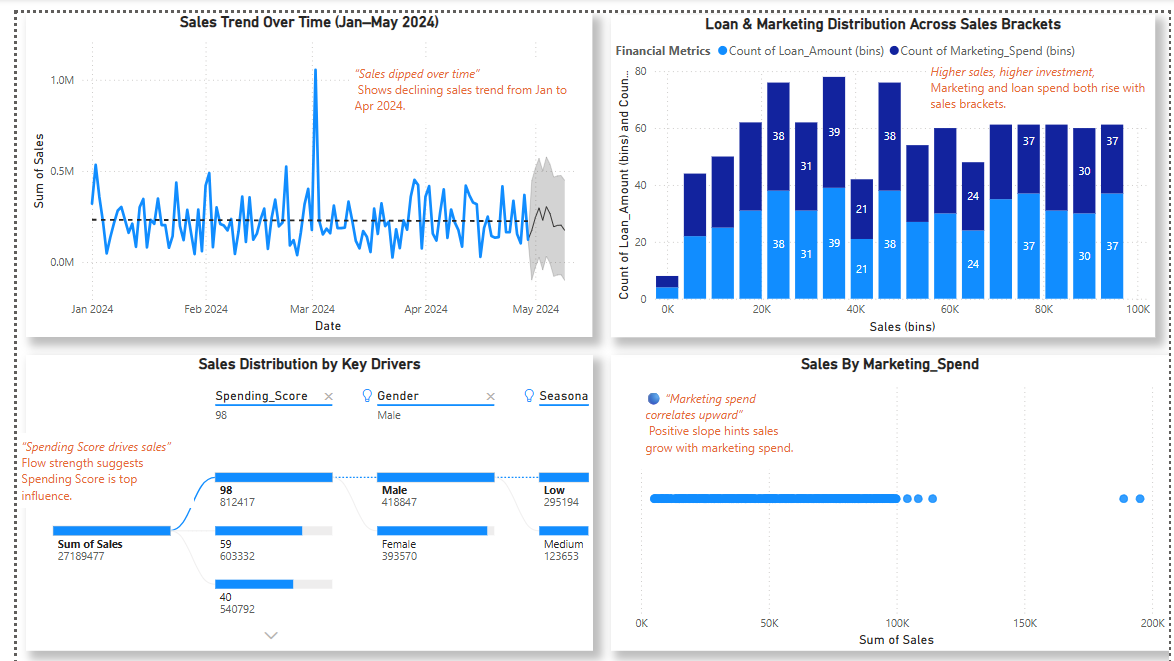
| **How aggressively AI searches for unusual data points directly proportional to sensitivity.** |  |
| --- | --- |
| **Churn rates vary by age and gender**, with males in the 25–34 and under-25 brackets showing notably higher churn. **Female churn is more stable**, but still spikes slightly in the 25–34 range across all visuals. |  |
| * **Males aged 65+** have the **highest churn rate (0.40)**. * Among **females**, churn is more stable and lower across age bands, with slight peaks in **Under 25** and **35–44**. * Overall, **churn rates are higher for males** in nearly every age category. | age has a stronger interaction effect.Gender's main effect is minimal. **young females or older males)** subtly influence churn direction. |

**Task 4: AI for Business Strategy and Risk Management**

**PowerBI Service Dashboard for Quick Insights**

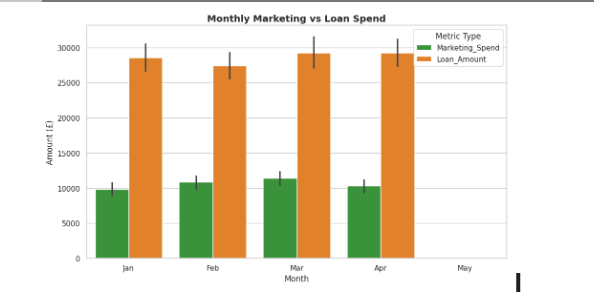
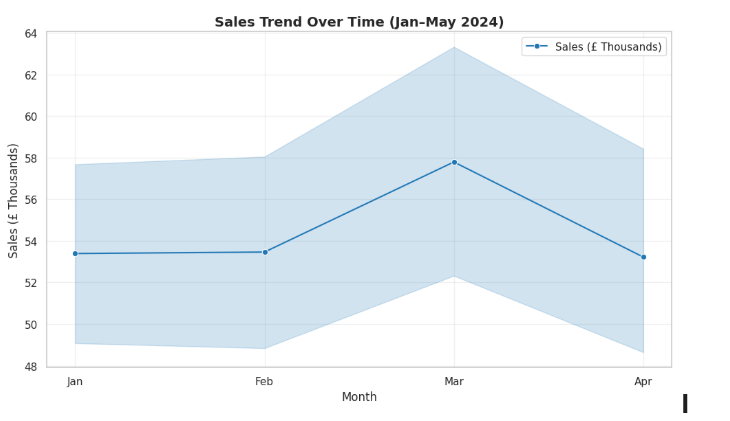
|  | **Power BI Desktop PPT Export** |
| --- | --- |

**Sales Performance & Key Growth Drivers (Jan–May 2024) PowerBI Predictions**

****

1. Sales Down Jan–Apr
2. Customers with Spending Score 98 Drives Sales
3. More Marketing, More Sales
4. Higher Spend, Higher Sales
5. Top: Males, Low Seasonality

**Sales Performance & Key Growth Drivers (Jan–May 2024) Python Predictions**

****

Insight Summary:

1. Sales declined steadily from Jan to Apr 2024; partial recovery in May.

2. Spending Score 98 customers contributed the most to total sales.

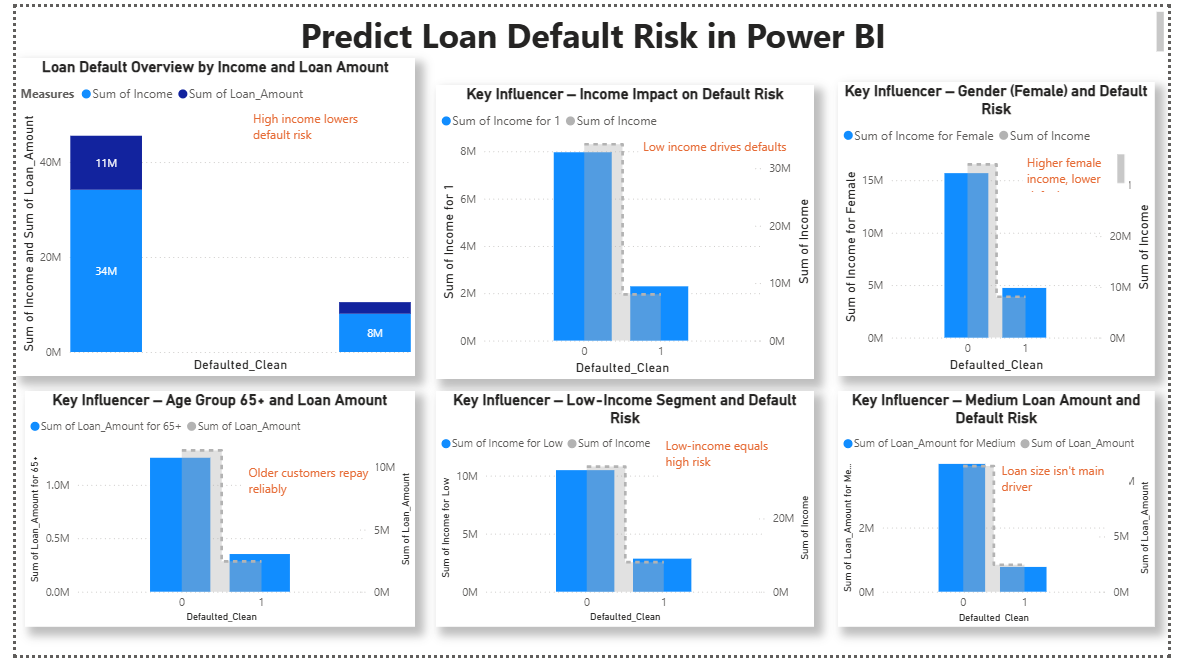
3. Marketing investment positively correlated with sales performance.

4. Loan and marketing spending increased with higher sales brackets.

5. Top performers: Male customers with low seasonality and high spending scores.

**Task 4: AI for Business Strategy and Risk Management using both Power BI and Python**

**Predict Loan Default Risk in Power BI**

** Key Insights List**

1. High income lowers default risk
2. Low income drives defaults
3. Higher female income, lower defaults
4. Older customers repay reliably
5. Low-income equals high risk
6. Loan size isn't main driver

**Predict Loan Default Risk By Python**

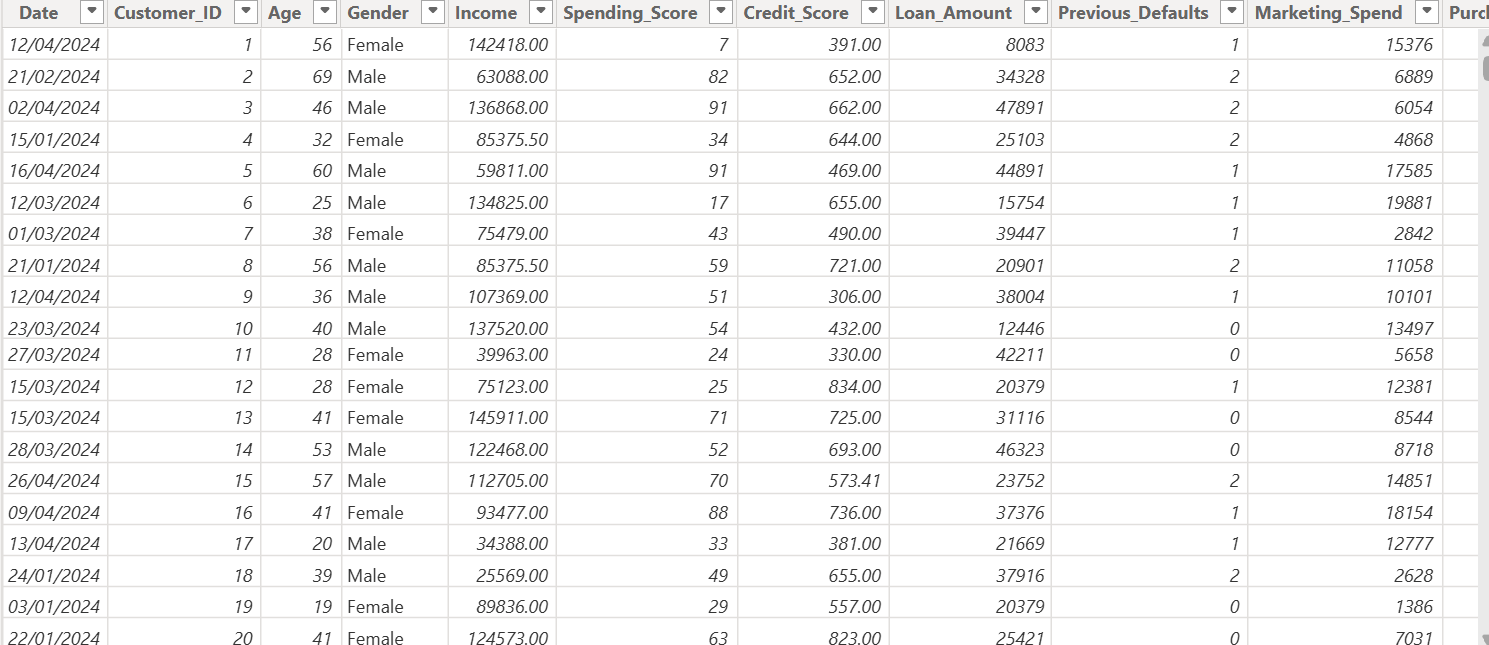
|  | **Key Insights**  Low income increases default risk  Moderate loan amounts matter  Credit score has minimal impact  Income is the strongest predictor  Balanced model captures defaulters well |
| --- | --- |

Deliverables

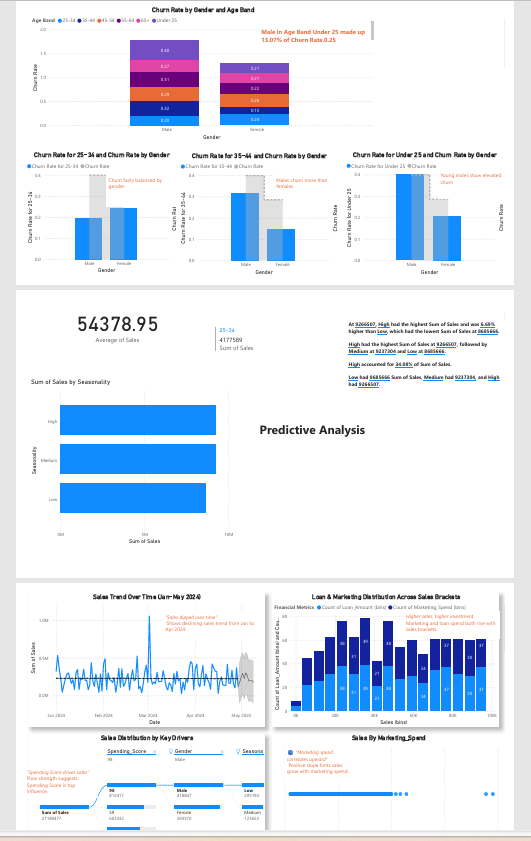
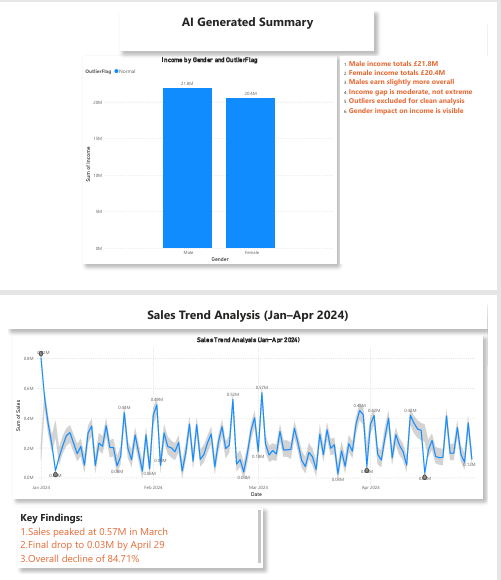
1.Python Script

<https://github.com/Srivalarmathi/Python/blob/main/Advanced_AI_Vs_Python.ipynb>

2.Cleaned Dataset



3.PowerBiGenerated PDF



### 5.Overall KeyInsights

1. **Income impacts default risk**
2. **March sales peaked sharply**
3. **Young males churn often**
4. **Spending Score boosts sales**
5. **Marketing spend drives growth**
6. **Target reliable loan segments**

### 6.Business Recommendations

1. **Boost seasonal marketing spend**
2. **Focus on high scorers**
3. **Retain churn-prone segments**
4. **Use predictive risk flags**
5. **Target reliable loan profiles**
6. **Plan resources by seasonality**

**Thank You Riya**

**Valarmathi Ganessin**