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.

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
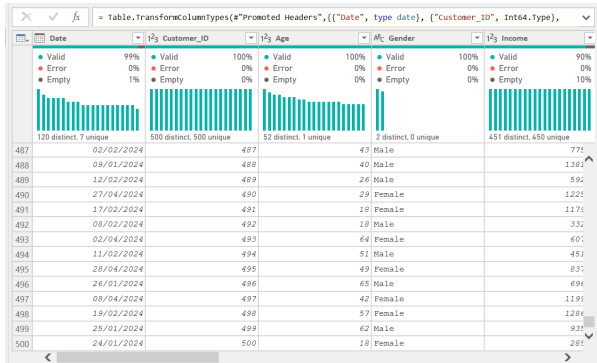**Interpret** – Understand model logic and variable impact
Strategize – Use insights for planning and forecasting
Communicate – Share results through visuals and summaries
Storytell – Present data-backed narratives visually and clearly

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	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

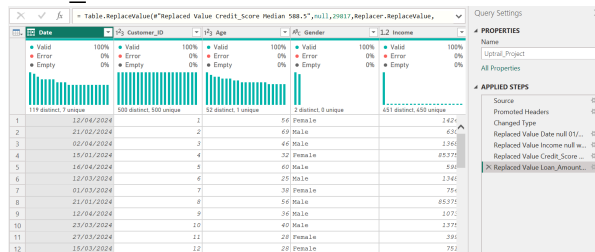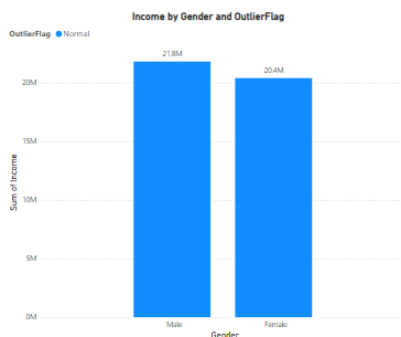|	
Age, Gender) ● **Errors**: % of entries that are incorrectly formatted (e.g., 1% error in Date) ● **Missing Values**: % of empty/null entries (e.g., 10% missing in Income)	Median: 54032.5 Standard Deviation: 27263.10646776021 Min: 5203 Max: 99835
2: **Handling Missing Values** Replaced column null 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**

Date	Customer_ID	Age	Gender	Income	Spending_Score	Credit_Score	Loan_Amount	Previous_Defaults	Marketing_Spend	Purch
12/04/2024	1	56	Female	142418	7	391	8083	1	15376	
21/02/2024	2	69	Male	63088	82	652	34328	2	6889	
02/04/2024	3	46	Male	136868	91	662	47891	2	6054	
15/01/2024	4	32	Female	85375.5	34	644	25103	2	4868	
16/04/2024	5	60	Male	59811	91	469	44891	1	17585	
12/03/2024	6	25	Male	134825	17	655	15754	1	19881	
01/03/2024	7	38	Female	75479	43	490	39447	1	2842	
21/01/2024	8	56	Male	85375.5	59	721	20901	2	11058	
12/04/2024	9	36	Male	107369	51	306	38004	1	10101	
23/03/2024	10	40	Male	137520	54	432	12446	0	13497	
27/03/2024	11	28	Female	39963	24	330	42211	0	5658	
15/03/2024	12	28	Female	75123	25	834	29817	1	12381	
15/03/2024	13	41	Female	145911	71	725	31116	0	8544	
28/03/2024	14	53	Male	122468	52	693	46323	0	8718	
26/04/2024	15	57	Male	112705	70	588.5	23752	2	14851	
09/04/2024	16	41	Female	93477	88	736	37376	1	18154	
13/04/2024	17	20	Male	34388	33	381	21669	1	12777	
24/01/2024	18	39	Male	25569	49	655	37916	2	2628	
03/01/2024	19	19	Female	89836	29	557	29817	0	1386	
22/01/2024	20	41	Female	124573	63	823	25421	0	7031	

Jptrail: Project (500 rows) Column: Date (119 distinct values)

Task 2: AI-Powered Data Visualisation and Storytelling

Sales Performance over time from **January to April 2024**. It highlights:

These anomalies are due to the aggregation of Dates. Sales peak in January.

- **Fluctuations in total sales** across dates
- **Peaks** like 0.91M in early January and 0.57M in March
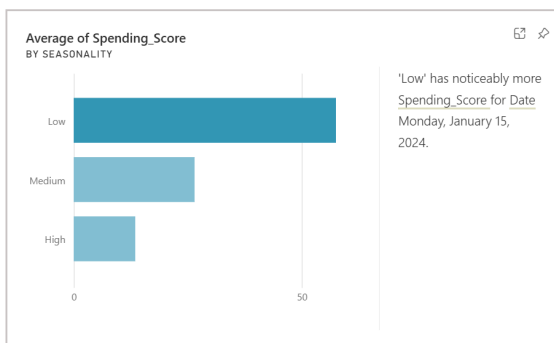- **Troughs** like 0.04M in February and 0.08M in April
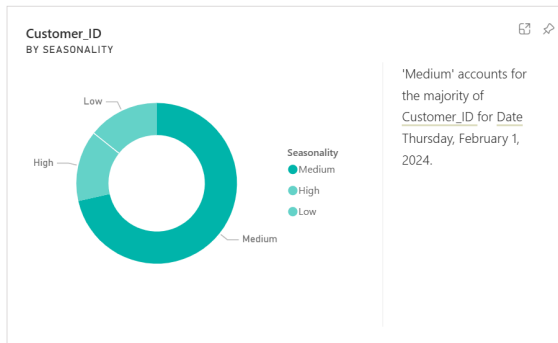
Shows February 1, 2024 has medium seasonality shopping.
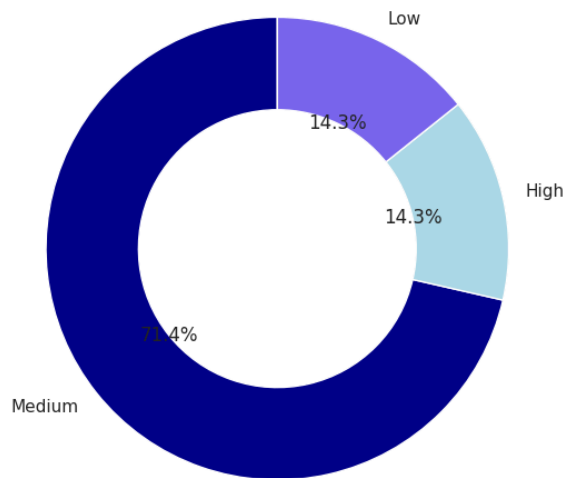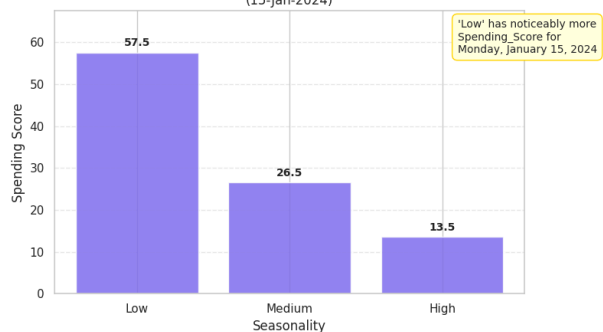

Thursday, February 1, 2024, with the *Medium* group clearly dominating.

is highlighting a surprising pattern in data for **Monday, January 15, 2024**:
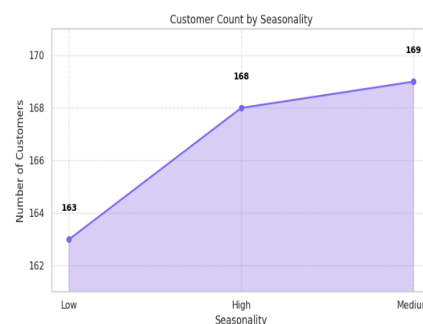

Low Seasonality Group having the Highest Spending Score on 15/01/2024

- 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.

Customer_ID
BY SEASONALITY

There is relatively even Customer_ID by Seasonality .
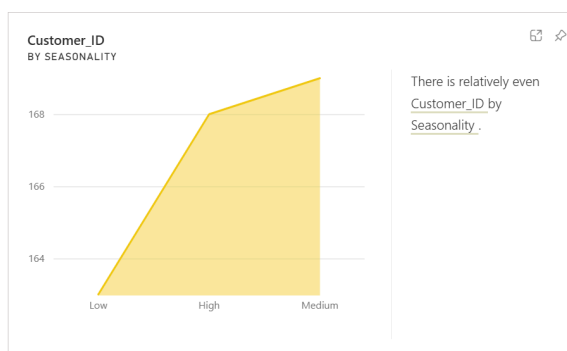
The number of customers is **fairly evenly distributed** across all three Seasonality groups

There's a **slight peak** in the *High Seasonality* group, but nothing extreme

Customer Count by Seasonality

There is relatively even distribution of customers by seasonality

Average of Credit_Score
BY SEASONALITY

There is relatively even Credit_Score by Seasonality .

Customers in the **Low Seasonality** group have the **highest average Credit Score**

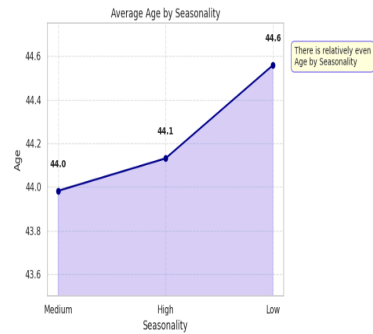Those in the **Medium Seasonality** group have the **lowest**

Average Credit Score by Seasonality

There is relatively even Credit_Score by Seasonality
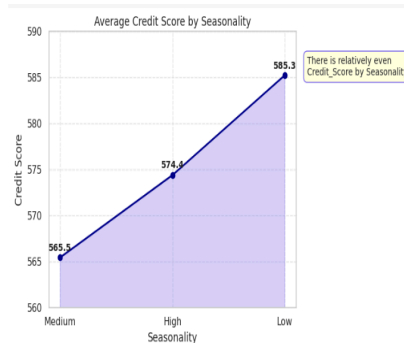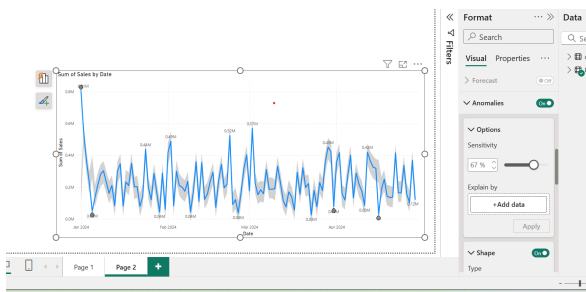
Relatively Even Credit_Score by Seasonality
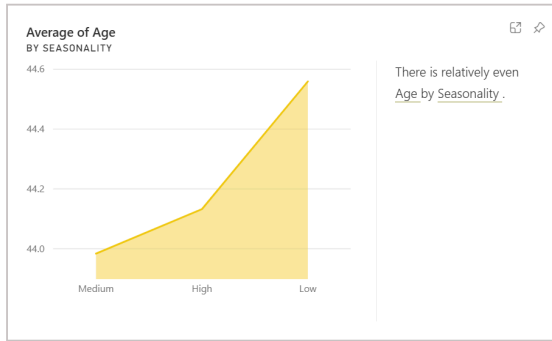
Customers in the **Low Seasonality** group are, on average, **older**

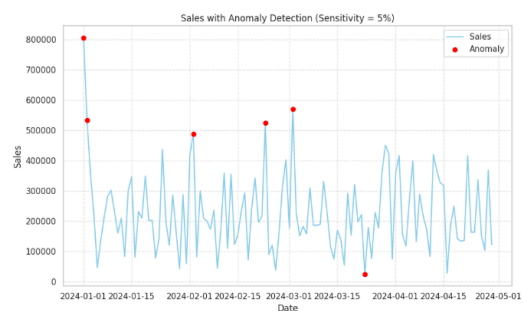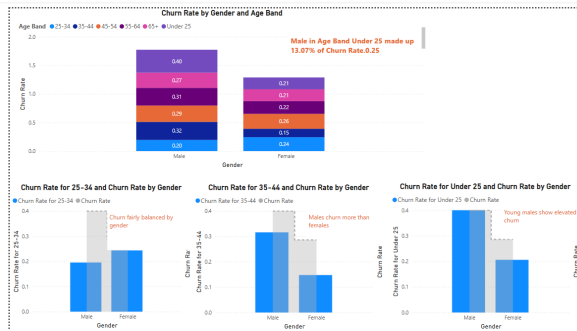- ○ Those in the **Medium Seasonality** group are the **youngest**

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.
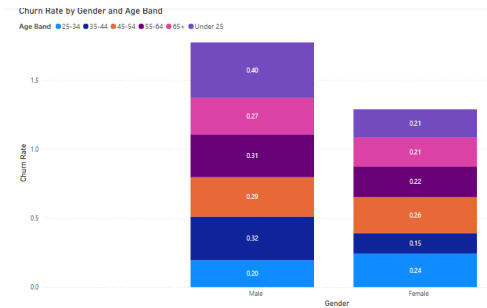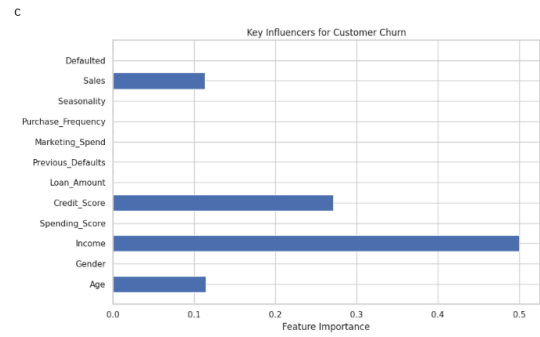
-
  Males aged 65+ have the **highest churn rate (0.40)**.
- Among **females**, churn is more stable and lower across age bands, with slight peaks in **Under 25** and **35–44**.
- Overall, **churn rates are higher for males** in nearly every age category.
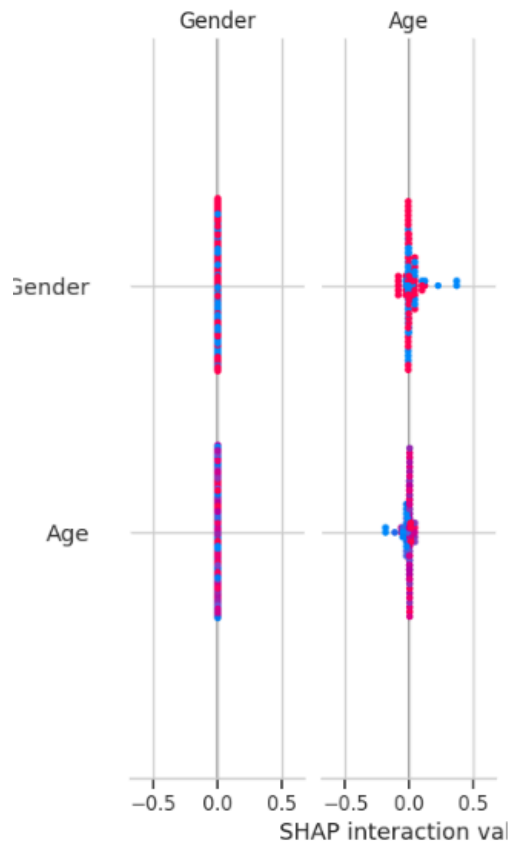
age has a stronger interaction effect. Gender's main effect is minimal. **young females or older males)** subtly influence churn direction.

Task 4: AI for Business Strategy and Risk Management
PowerBI Service Dashboard for Quick Insights
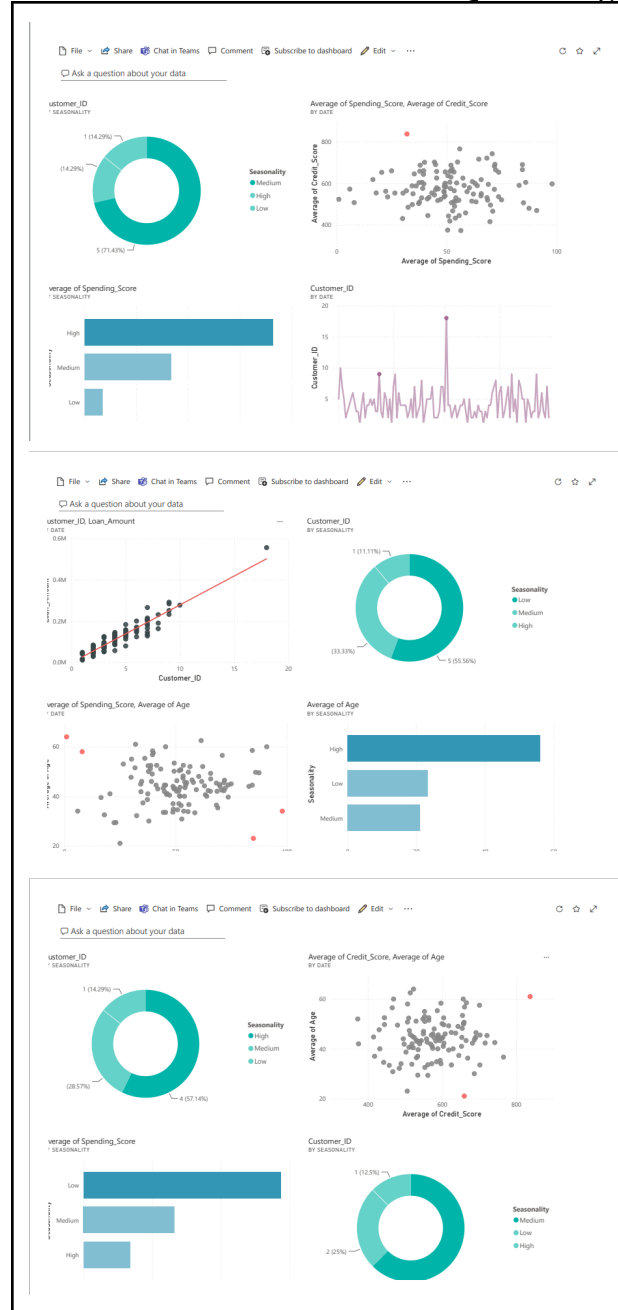
Power BI Desktop PPT Export
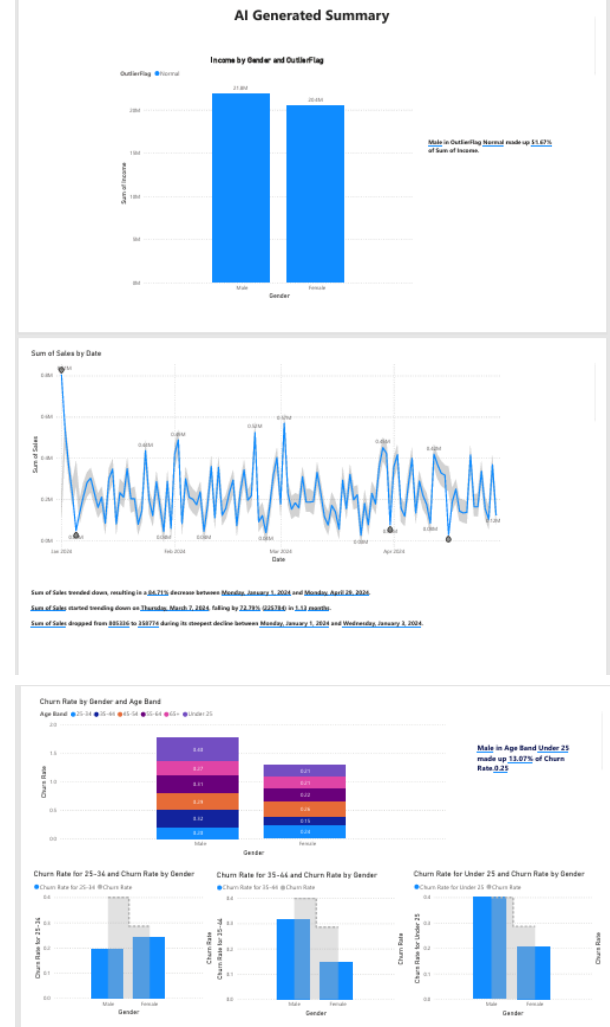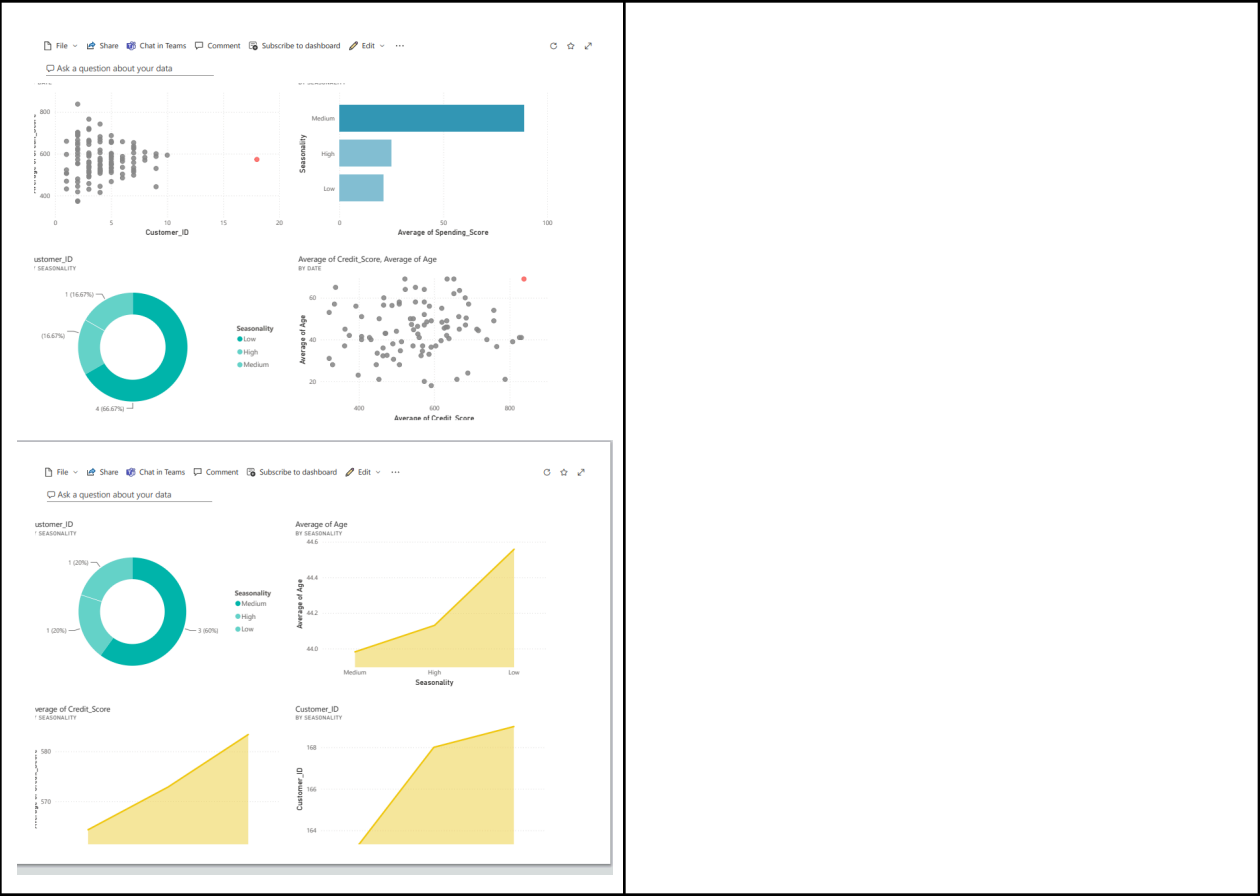
Sales Performance & Key Growth Drivers (Jan–May 2024) PowerBI Predictions
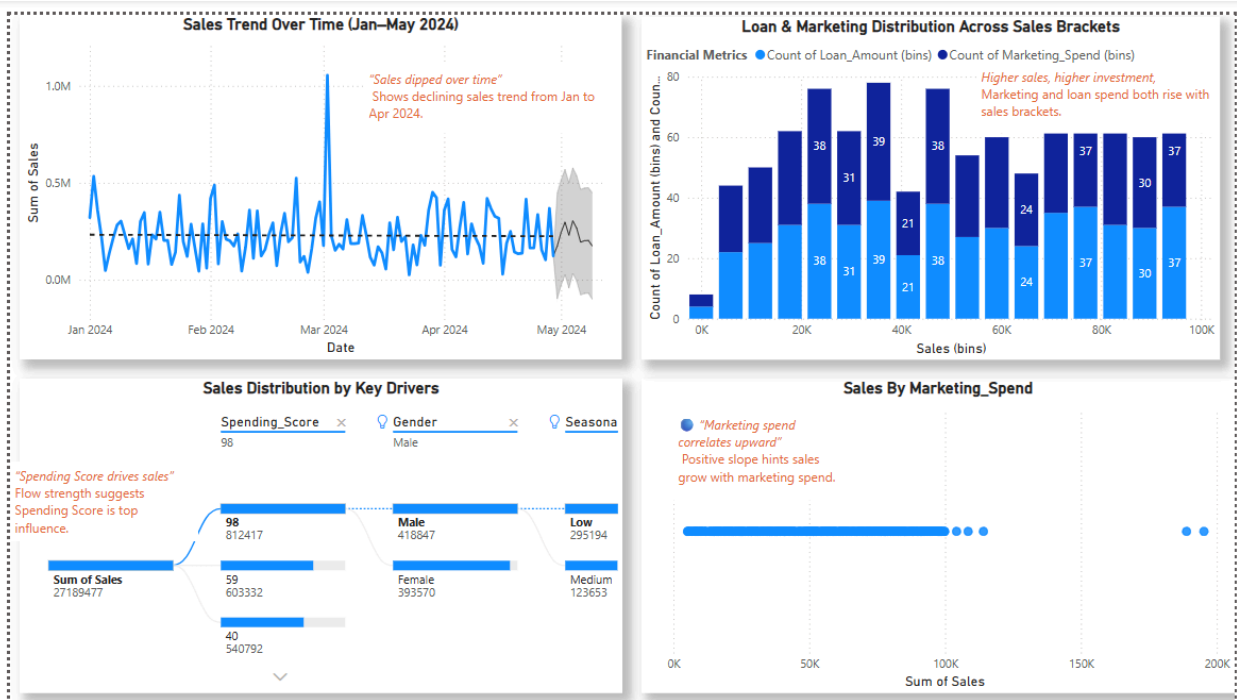
1. Sales Down Jan–Apr
2. Customers with Spending Score 98 Drives Sales

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
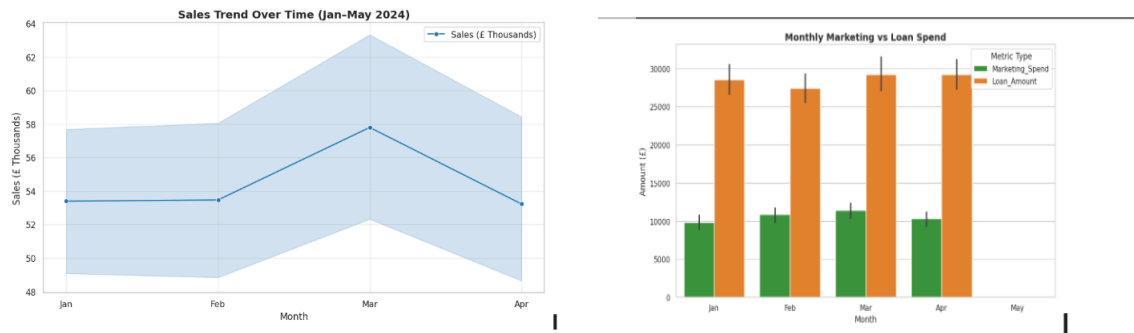
Insight Summary:
1. Sales declined steadily from Jan to Apr 2024; partial recovery in May.
2. Spending Score 98 customers contributed the most to total sales.
3. Marketing investment positively correlated with sales performance.
4. Loan and marketing spending increased with higher sales brackets.
5. 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

Predict Loan Default Risk in Power BI
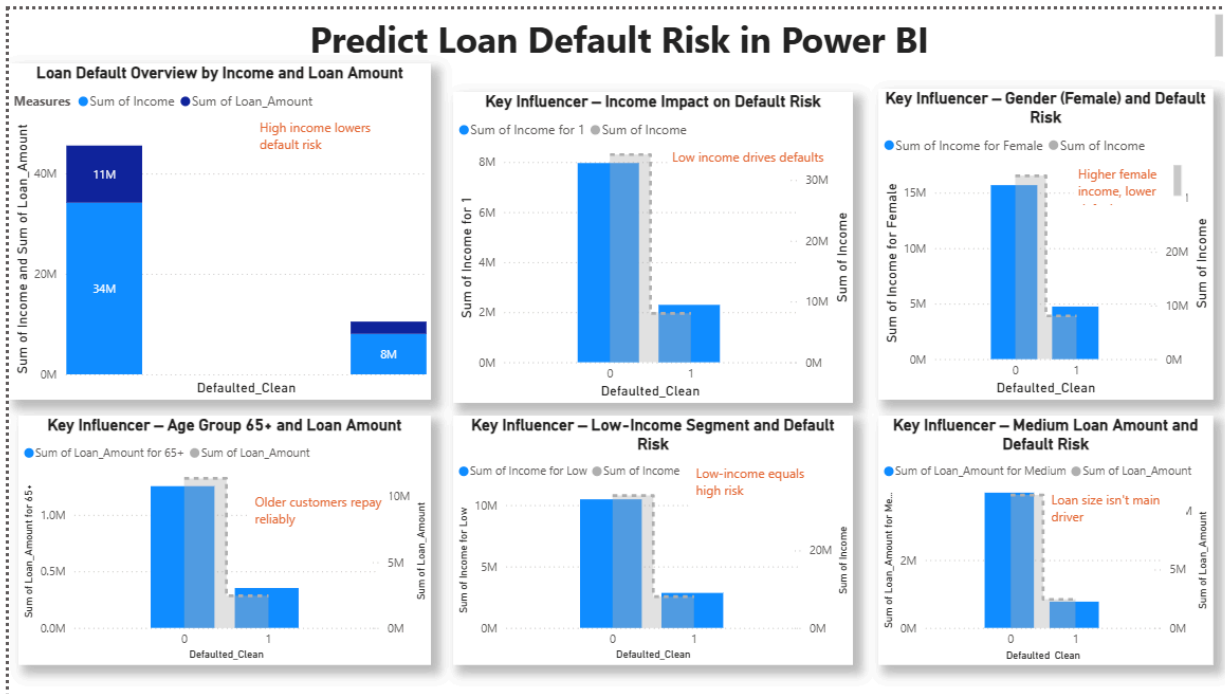
Key Insights List

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
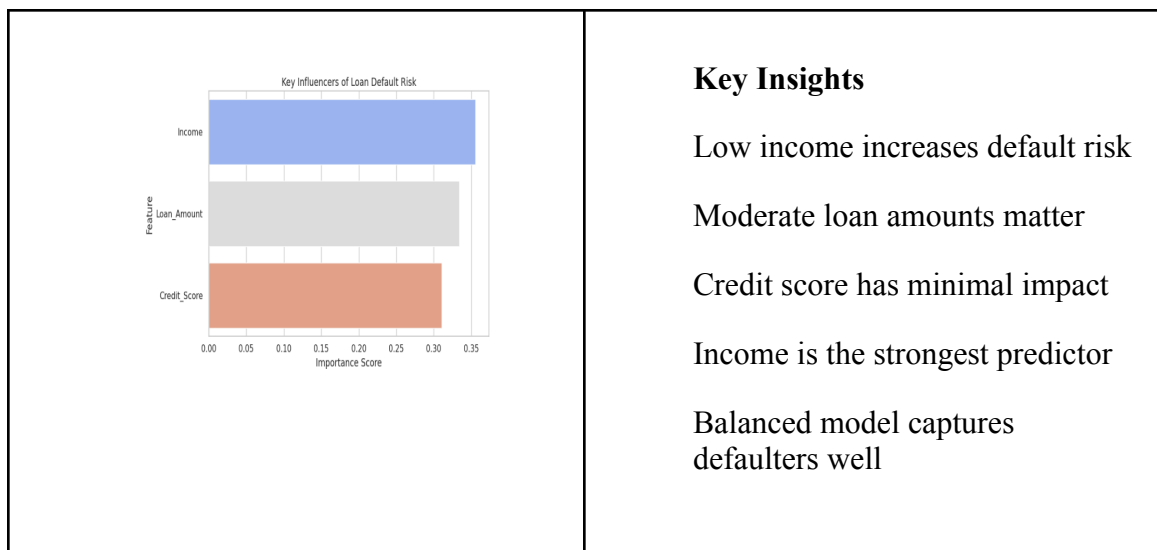
Deliverables

1.Python Script

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

2.Cleaned Dataset

Date	Customer_ID	Age	Gender	Income	Spending_Score	Credit_Score	Loan_Amount	Previous_Defaults	Marketing_Spend	Purc
12/04/2024	1	56	Female	142418.00	7	391.00	8083	1	15376	
21/02/2024	2	69	Male	63088.00	82	652.00	34328	2	6889	
02/04/2024	3	46	Male	136868.00	91	662.00	47891	2	6054	
15/01/2024	4	32	Female	85375.50	34	644.00	25103	2	4868	
16/04/2024	5	60	Male	59811.00	91	469.00	44891	1	17585	
12/03/2024	6	25	Male	134825.00	17	655.00	15754	1	19881	
01/03/2024	7	38	Female	75479.00	43	490.00	39447	1	2842	
21/01/2024	8	56	Male	85375.50	59	721.00	20901	2	11058	
12/04/2024	9	36	Male	107369.00	51	306.00	38004	1	10101	
23/03/2024	10	40	Male	137520.00	54	432.00	12446	0	13497	
27/03/2024	11	28	Female	39963.00	24	330.00	42211	0	5658	
15/03/2024	12	28	Female	75123.00	25	834.00	20379	1	12381	
15/03/2024	13	41	Female	145911.00	71	725.00	31116	0	8544	
28/03/2024	14	53	Male	122468.00	52	693.00	46323	0	8718	
26/04/2024	15	57	Male	112705.00	70	573.41	23752	2	14851	
09/04/2024	16	41	Female	93477.00	88	736.00	37376	1	18154	
13/04/2024	17	20	Male	34388.00	33	381.00	21669	1	12777	
24/01/2024	18	39	Male	25569.00	49	655.00	37916	2	2628	
03/01/2024	19	19	Female	89836.00	29	557.00	20379	0	1386	
22/01/2024	20	41	Female	124573.00	63	823.00	25421	0	7031	

3.PowerBiGenerated PDF

AI Generated Summary

Income by Gender and OutlierFlag

OutlierFlag ● Normal
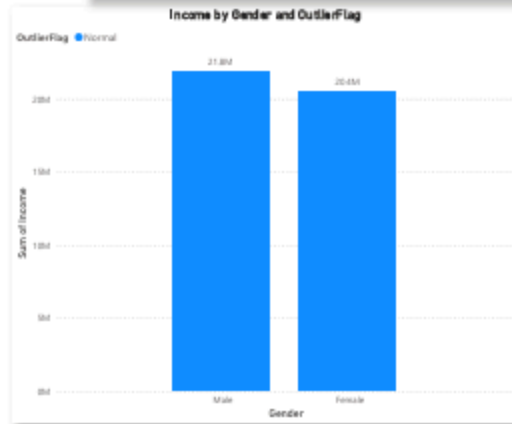
1. **Male income totals £21.8M**
2. **Female income totals £20.4M**
3. **Males earn slightly more overall**
4. **Income gap is moderate, not extreme**
5. **Outliers excluded for clean analysis**
6. **Gender impact on income is visible**

Sales Trend Analysis (Jan–Apr 2024)

Key Findings:
1. Sales peaked at 0.57M in March
2. Final drop to 0.03M by April 29
3. Overall decline of 84.71%

Churn Rate by Gender and Age Band

Age Band ● 25-34 ● 35-44 ● 45-54 ● 55-64 ● 65+ ● Under 25

Churn Rate

Male in Age Band Under 25 made up 13.07% of Churn Rate.0.25

Male bars: 0.48, 0.27, 0.31, 0.29, 0.32, 0.28
Female bars: 0.21, 0.21, 0.22, 0.26, 0.15, 0.28

Gender

Churn Rate for 25-34 and Churn Rate by Gender
● Churn Rate for 25-34 ● Churn Rate

Churn fairly balanced by gender

Churn Rate for 25-34 / Gender

Churn Rate for 35-44 and Churn Rate by Gender
● Churn Rate for 35-44 ● Churn Rate

Males churn more than females

Churn Rate for 35-44 / Gender

Churn Rate for Under 25 and Churn Rate by Gender
● Churn Rate for Under 25 ● Churn Rate

Young males show elevated churn

Churn Rate / Gender

54378.95
Average of Sales

25-34
4177589
Sum of Sales

At 9266507, High had the highest Sum of Sales and was 6.69% higher than Low, which had the lowest Sum of Sales at 8685666.

High had the highest Sum of Sales at 9266507, followed by Medium at 9237304 and Low at 8685666.

High accounted for 34.00% of Sum of Sales.

Low had 8685666 Sum of Sales, Medium had 9237304, and High had 9266507.

Sum of Sales by Seasonality

Seasonality: High, Medium, Low
Sum of Sales (0M – 10M)

Predictive Analysis

Sales Trend Over Time (Jan-May 2024)

"Sales dipped over time"
Shows declining sales trend from Jan to Apr 2024.

Sum of Sales (0.0M – 1.0M)
Date: Jan 2024, Feb 2024, Mar 2024, Apr 2024, May 2024

Loan & Marketing Distribution Across Sales Brackets

Financial Metrics ● Count of Loan_Amount (bins) ● Count of Marketing_Spend (bins)

Higher sales, higher investment; Marketing and loan spend both rise with sales brackets.

Count of Loan_Amount (bins) and Cou...
Sales (bins): 0K, 20K, 40K, 60K, 80K, 100K

Sales Distribution by Key Drivers

Spending_Score × Gender × Seasona...
98   Male

"Spending Score drives sales"
Flow strength suggests Spending Score is top influence.

Sum of Sales 98 912417   Male 418687   Low 291150
27308477   59 683252   Female 293570   Medium 12165.0

Sales By Marketing_Spend

● *"Marketing spend correlates upward"*
Positive slope hints sales grow with marketing spend.

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

7. Boost seasonal marketing spend
8. Focus on high scorers
9. Retain churn-prone segments
10. Use predictive risk flags
11. Target reliable loan profiles
12. Plan resources by seasonality

Thank You   **Riya**

Valarmathi Ganessin