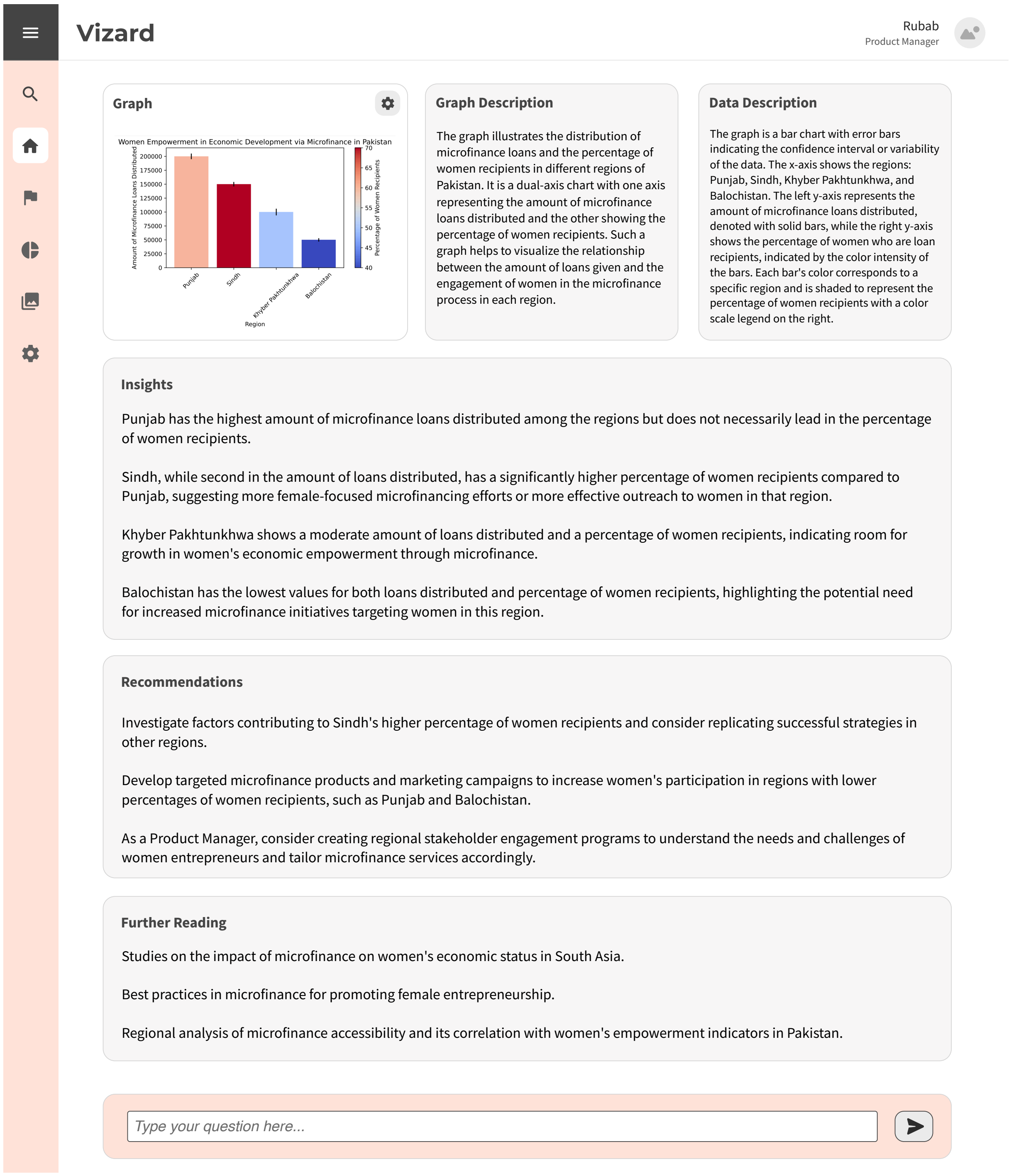
APPENDIX


Vizard
Rubab
Product Manager
Graph
Women Empowerment in Economic Development via Microfinance in Pakistan
Amount of Microfinance Loans Distributed
200000
175000
150000
125000
100000
75000
50000
25000
0
Punjab
Sindh
Khyber Pakhtunkhwa
Balochistan
Region
Percentage of Women Recipients
70
65
60
55
50
45
40

Graph Description

The graph illustrates the distribution of microfinance loans and the percentage of women recipients in different regions of Pakistan. It is a dual-axis chart with one axis representing the amount of microfinance loans distributed and the other showing the percentage of women recipients. Such a graph helps to visualize the relationship between the amount of loans given and the engagement of women in the microfinance process in each region.

Data Description

The graph is a bar chart with error bars indicating the confidence interval or variability of the data. The x-axis shows the regions: Punjab, Sindh, Khyber Pakhtunkhwa, and Balochistan. The left y-axis represents the amount of microfinance loans distributed, denoted with solid bars, while the right y-axis shows the percentage of women who are loan recipients, indicated by the color intensity of the bars. Each bar's color corresponds to a specific region and is shaded to represent the percentage of women recipients with a color scale legend on the right.

Insights

Punjab has the highest amount of microfinance loans distributed among the regions but does not necessarily lead in the percentage of women recipients.

Sindh, while second in the amount of loans distributed, has a significantly higher percentage of women recipients compared to Punjab, suggesting more female-focused microfinancing efforts or more effective outreach to women in that region.

Khyber Pakhtunkhwa shows a moderate amount of loans distributed and a percentage of women recipients, indicating room for growth in women's economic empowerment through microfinance.

Balochistan has the lowest values for both loans distributed and percentage of women recipients, highlighting the potential need for increased microfinance initiatives targeting women in this region.

Recommendations

Investigate factors contributing to Sindh's higher percentage of women recipients and consider replicating successful strategies in other regions.

Develop targeted microfinance products and marketing campaigns to increase women's participation in regions with lower percentages of women recipients, such as Punjab and Balochistan.

As a Product Manager, consider creating regional stakeholder engagement programs to understand the needs and challenges of women entrepreneurs and tailor microfinance services accordingly.

Further Reading

Studies on the impact of microfinance on women's economic status in South Asia.

Best practices in microfinance for promoting female entrepreneurship.

Regional analysis of microfinance accessibility and its correlation with women's empowerment indicators in Pakistan.

Type your question here...

Figure 6: A visualization of our dashboard interface. Each section corresponds to a different level of information being extracted from the graph. Users can upload an image of their choice to be analyzed or launch VIZARD through its integration in their dashboard software.

Table 3: LLM instructions tailored to the user's literacy level. Each higher level assumes knowledge of previous levels.

Level	Definition
1	**The user does not know anything about data visualizations such as graphs and charts.** ↪ The output should explain the fundamental concepts of the visualization in question in language that is easy to understand. Name the type of visualization, explain its components, and explain how and why it is used. The user cannot consume technical insights and recommendations. The user would really like to enhance their ability to remember this type of chart. Give the user some limited insights and recommendations that can be drawn from the visualization by applying their domain knowledge in an intuitive way. Examples should be added as well according to their domain and role.
2	**The user knows about different kinds of visualizations but finds it difficult to build relationships between the variables given in them.** ↪ The output should focus less on explaining the graph in question and more on why it's used here and how to interpret the relationships between the variables. The user would really like to learn how to build relationships between different variables. The output should take into account their domain and role.
3	**The user can build relationships between two variables but finds it difficult to build relationship between more variables.** ↪ The output of all the sections should be balanced with knowledge. The output should take into account the user's domain and role.
4	**The user can build relationships between more than two variables but finds it difficult to apply their domain knowledge to the visualization.** ↪ The output should be focused on incorporating domain knowledge into the graph. The output should take into account the user's domain and role.
5	**The user can apply domain knowledge and derive insights and recommendations from the visualizations.** ↪ The output should include inter- and intra-domain information, market trends, and questions for the user to think about in the insights and recommendations sections. The further readings section should be technically resourceful. The output should take into account the user's domain and role.

Table 4: Plot Difficulty Levels. The following instructions are substituted in the prompt as `"The plots should be of {level}"`. The full idea generation prompt can be found in Figure 7.

Level	Definition
1	`"low complexity and use a few graph elements from the list."`
2	`"medium complexity and use some graph elements from the list."`
3	`"high complexity and use most graph elements from the list."`

Table 5: Question Difficulty Levels. The following instructions are substituted in the prompt as `"The questions should focus on {level}"`. The full question generation prompt can be found in Figure 9.

Level	Definition
1	`"the graphical elements of the plot and what they represent for the data."`
2	`"the trends that can be observed in the data from looking at the plot."`
3	`"the insights that can be drawn from the trends in the data and the actions that can be taken based on them."`

```
Task: Your job is to help generate ideas for graphical plots on a certain topic using different chart styles and variables and their
combinations. The plots should be of {p_level}. Return the output as a list of JSON objects of key-value pairs, one for each graph.

Instructions: Generate ideas for plots on the topic of {plot_topic} in the following format:

1. Key: "chart_type"
   Value: The chart type, for example "line chart", "bar chart", "scatter plot", "violin plot", "heatmap", "box plot", "contour
   plot", "bubble chart", "histogram", or other type of plot.

2. Key: "variables"
   Value: A list of variables involved in the plot, for example "age", "gender", "height", etc.

3. Key: "x-axis"
   Value: The variable that should be plotted on the x-axis. Leave blank if not applicable to this chart type.

4. Key: "y-axis"
   Value: The variable that should be plotted on the y-axis. Leave blank if not applicable to this chart type.

5. Key: "color"
   Value: The variable that should be used to color the data points. Leave blank if not applicable to this chart type.

6. Key: "style"
   Value: The variable that should be used to represent the style of the categories (e.g. solid, dashed, dotted, shading).
   Leave blank if not applicable to this chart type.

7. Key: "label"
   Value: The variable that should be used to label the data points. Leave blank if not applicable to this chart type.

8. Key: "sizes"
   Value: The variable that should be used to define the sizes of the data points. Leave blank if not applicable to this chart
   type.

9. Key: "error_bars"
   Value: True or False. Leave blank if not applicable to this chart type.

10. Key: "instructions"
    Value: Additional instructions in natural language that are required to accurately specify the chart appearance.

Do not return anything except the list of JSON objects of key-value pairs as output.
```

Figure 7: The plot idea generation prompt. The variables in red are set based on the desired plot difficulty level and topic of interest. The definitions of the different plot difficulty levels can be found in Table 4 and the plot topic is provided in the form of a natural language description, e.g. "women empowerment". An example output using this prompt is shown in Figure 11.

```
Task: Your job is to generate a plot according to the instructions and JSON schema
provided. If you do not have access to the data required for it, you should generate
hypothetical data that seems plausible and reflects realistic trends.

Instructions: {instructions}

Schema: {schema}

Do not return anything except the Python code for the plot as output.
```

Figure 8: The plot code generation prompt. The instructions provided are those produced by the LLM in the idea generation stage in Figure 7 and the schema is the rest of the plot description provided in that response. An example output using this prompt is shown in Figure 12.

```
Task: Your job is to generate a multiple choice question to test a user's data literacy
level. You will be provided the description and Python code for a data plot and will be
asked to generate a question followed by four answer options, one per line, with the
correct option marked with a star (*) at the beginning. The question should focus on
{q_level}.

Instructions: {instructions}

Code: {code}

Do not return anything except the question and the answer options as output.
```

Figure 9: The question generation prompt. The instructions and code provided are those obtained by prompting the LLM in the plot code generation stage detailed in Figure 8. The definitions of the different question difficulty levels can be found in Table 5.

```
Task: Your task is to generate helpful educational content around the graph image that is uploaded. The content should be tailored to the
individual using them, provided to you as input. The output should be content around the graph broken down into multiple sections.

Instructions: Generate the graph content for a user who has following characteristics. The detail level of the output should be dictated by
the user's ability.

1. The user understands {language} as their preferred language.
2. The user has a role of {user_role} at their workplace, and {domain} is the industry they work in.
3. {dl_level}

The output should have the following schema:

1. Key: "graph_description"
   Value: The description of the graph shown in the image.

2. Key: "data_description"
   Value: The description of the labels, axis, colors, legends, data points, error bars.

3. Key: "insights"
   Value: Bullet-point-wise insights that can be drawn from it given {domain} as domain and {user_role} as the user's role.

4. Key: "recommendations"
   Value: The actionable information tailored to a user with a role of {user_role} in {domain} domain.

5. Key: "further_links"
   Value: Further references to studies and topics relevant to the graph and trends.

Do not return anything except the list of JSON objects of key-value pairs as output.
```

Figure 10: The educational content generation prompt. The variables in red are defined according to the user's preferences. Table 3 shows a mapping from the user's data visualisation literacy level to the instructions provided to the LLM to generate educational content tailored for them.

```json
{
    "topic": "women empowerment in economic development and making them inclusive using
microfinance especially in pakistan",
    "plot_level": 2,
    "question_level": 1,
    "chart_type": "bar chart",
    "variables": "['region', 'amount of microfinance loans distributed', 'percentage of
women recipients']",
    "x-axis": "region",
    "y-axis": "amount of microfinance loans distributed",
    "color": "percentage of women recipients",
    "style": null,
    "label": null,
    "sizes": null,
    "error_bars": true,
    "instructions": "Create a bar chart where each bar represents a region in Pakistan and
is colored based on the percentage of women recipients. Add error bars to represent
variability."
}
```

Figure 11: A JSON schema produced by GPT-4 using the prompt in Figure 7. This schema was used to generate the code in Figure 12.

```python
import matplotlib.pyplot as plt
import numpy as np

# Hypothetical data for regions in Pakistan
regions = ['Punjab', 'Sindh', 'Khyber Pakhtunkhwa', 'Balochistan']
amount_loans = np.array([200000, 150000, 100000, 50000])
percentage_women_recipients = np.array([60, 70, 50, 40])
error = np.array([5000, 4000, 6000, 3000])  # Hypothetical variability in the distribution of loans

# Normalize the percentage for color mapping
norm = plt.Normalize(percentage_women_recipients.min(), percentage_women_recipients.max())
sm = plt.cm.ScalarMappable(cmap="coolwarm", norm=norm)
sm.set_array([])

# Creating the bar plot
plt.bar(regions, amount_loans, yerr=error, color=sm.to_rgba(percentage_women_recipients))
plt.colorbar(sm, label='Percentage of Women Recipients')

plt.xlabel('Region')
plt.ylabel('Amount of Microfinance Loans Distributed')
plt.title('Women Empowerment in Economic Development via Microfinance in Pakistan')
plt.xticks(rotation=45)
plt.tight_layout()

plt.show()
```

Figure 12: Python code generated by GPT-4 for the JSON schema defined in Figure 11. This code was used to generate the plot in Figure 3b.

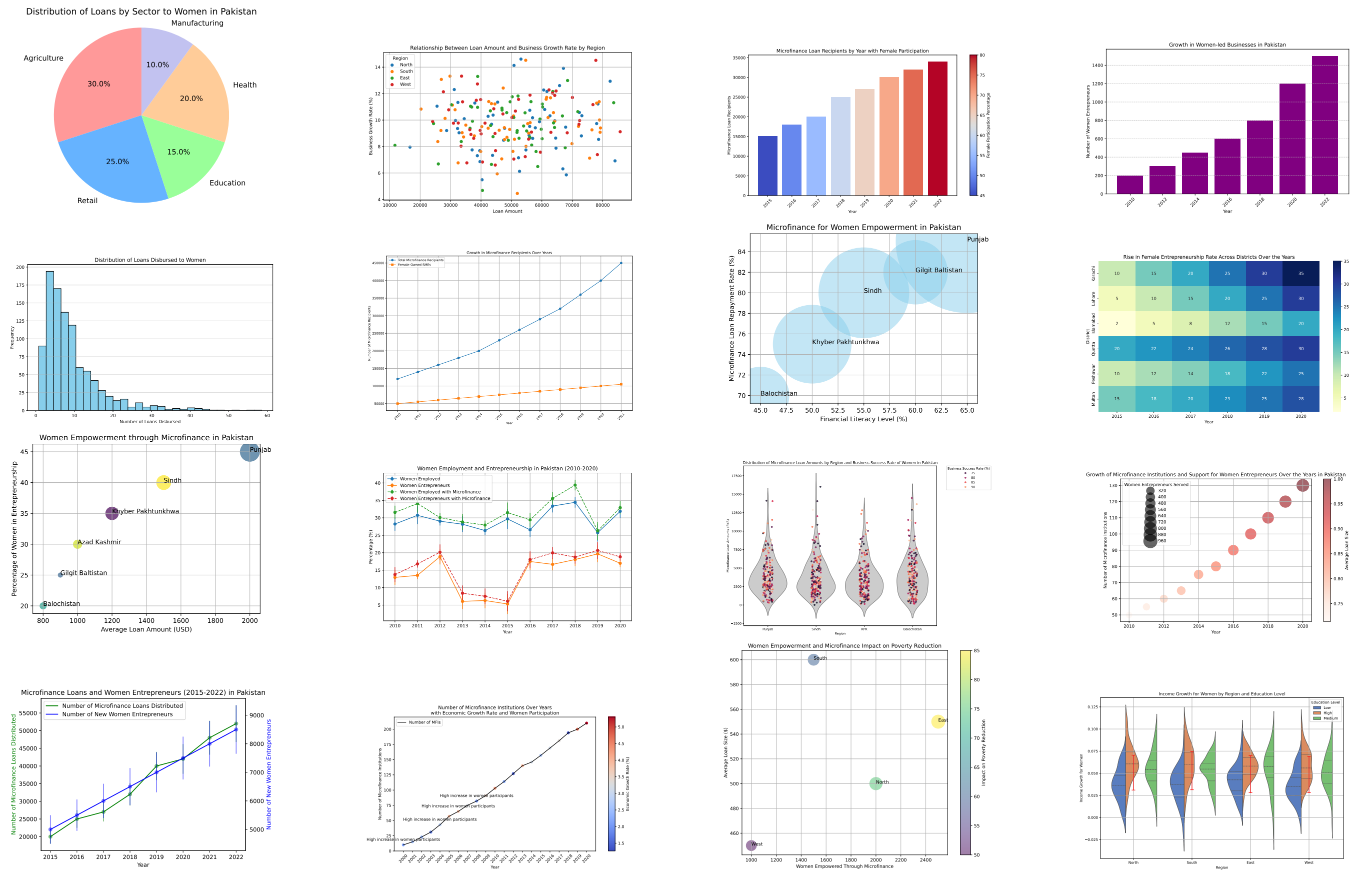


Figure 13: Additional examples showcasing the variation in type and complexity of data visualizations that GPT-4 is able to generate. All plots shown here are produced as-is with no human intervention.