



HarrisX:

AI-powered Research Chatbot for Smarter Decisions

Table of contents



- 01** Introduction and Problem Statements
- 02** AI Solution: HarrisX Research Chatbot
- 03** Data Analysis and Pipeline Overview
- 04** Demonstration of Effectiveness
- 05** Financial Benefit & ROI
- 06** Risks & Mitigations & Impacts

Market Research Firm HarrisX Seeks a More Efficient and Intelligent Research Workflow



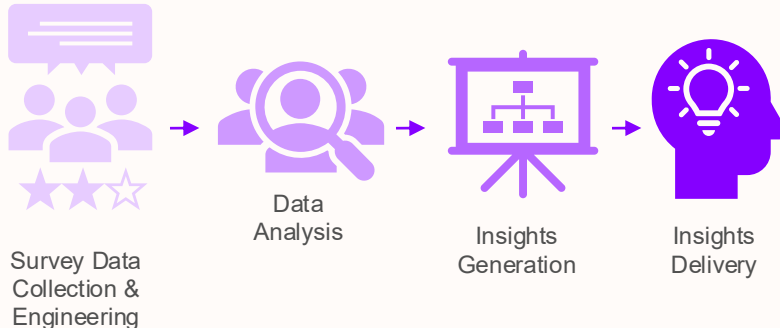
Company Intro

A tech-enabled **market research and analytics firm**, with clients from Fortune 500 companies

Providing **data-driven solutions** based on **opinion polling and consumer insights** worldwide

Current Workflow of providing services

causes **Issues** on:



1. Time-consuming and Costly Procedures
2. Strong Reliance on Human Input
3. Inconvenience for Non-technical Users

The HarrisX Research Chatbot (RCH): Leveraging AI to Boost Operational Efficiency

Example of user cases



"What % of Americans are cutting back on car purchases in Q1 2025?"

RCH interprets the question → pulls the right data → provides SQL query and key findings → shows a chart

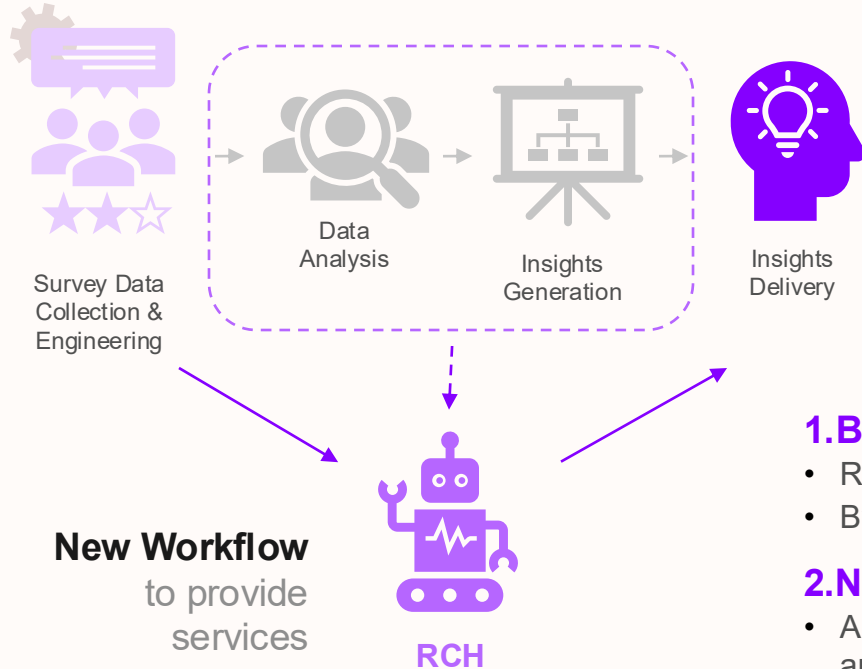
Expected Benefits of HarrisX Research Chatbot

1. Built for Speed & Accessibility

- Reduces reliance on human input and analyst bandwidth
- Bypasses manual data pulls and dashboard limitations

2. Natural Language Interface

- Allows typing plain-language questions to get data driven analysis and insights
- Empowers non-technical users to explore complex datasets



Chatbot Demo Built on Real-World Customer Survey Data

1. Your Gender	2. Your Age	3. Are you currently....?	4. What is your annual income?	5. How often do you visit Starbucks?
Female	From 20 to 29	Student	Less than RM25,000	Rarely
Female	From 20 to 29	Student	Less than RM25,000	Rarely
Male	From 20 to 29	Employed	Less than RM25,000	Monthly
Female	From 20 to 29	Student	Less than RM25,000	Rarely
Male	From 20 to 29	Student	Less than RM25,000	Monthly

Dataset Summary

The dataset from Kaggle contains 21 columns and 121 records, capturing responses from a customer satisfaction survey focused on Starbucks visitors.

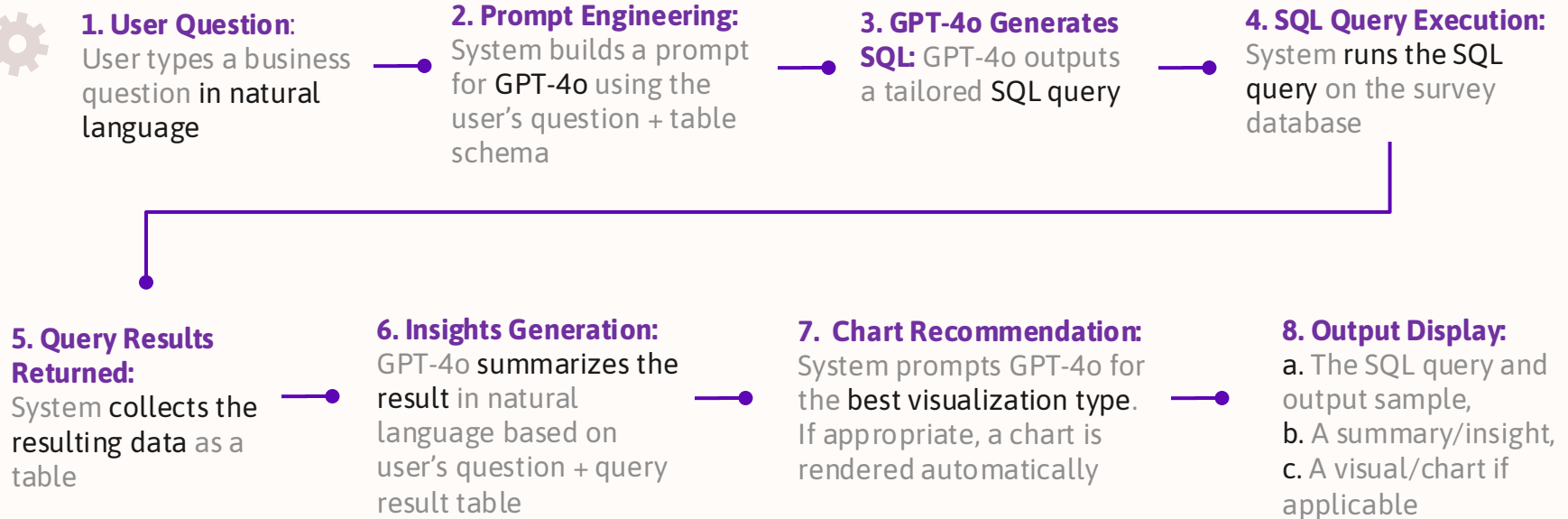
Survey Question Types

Survey fields span demographics, customer behavior, spending habits, satisfaction and experience ratings (on a 1–5 scale), loyalty/retention, and marketing channel awareness.


Data Structure & Quality

Data types include categorical, numerical, boolean, and datetime fields. Most columns have no missing values, ensuring high data quality for reliable analysis.

Unleashing the Power of Generative AI in Chatbot Development



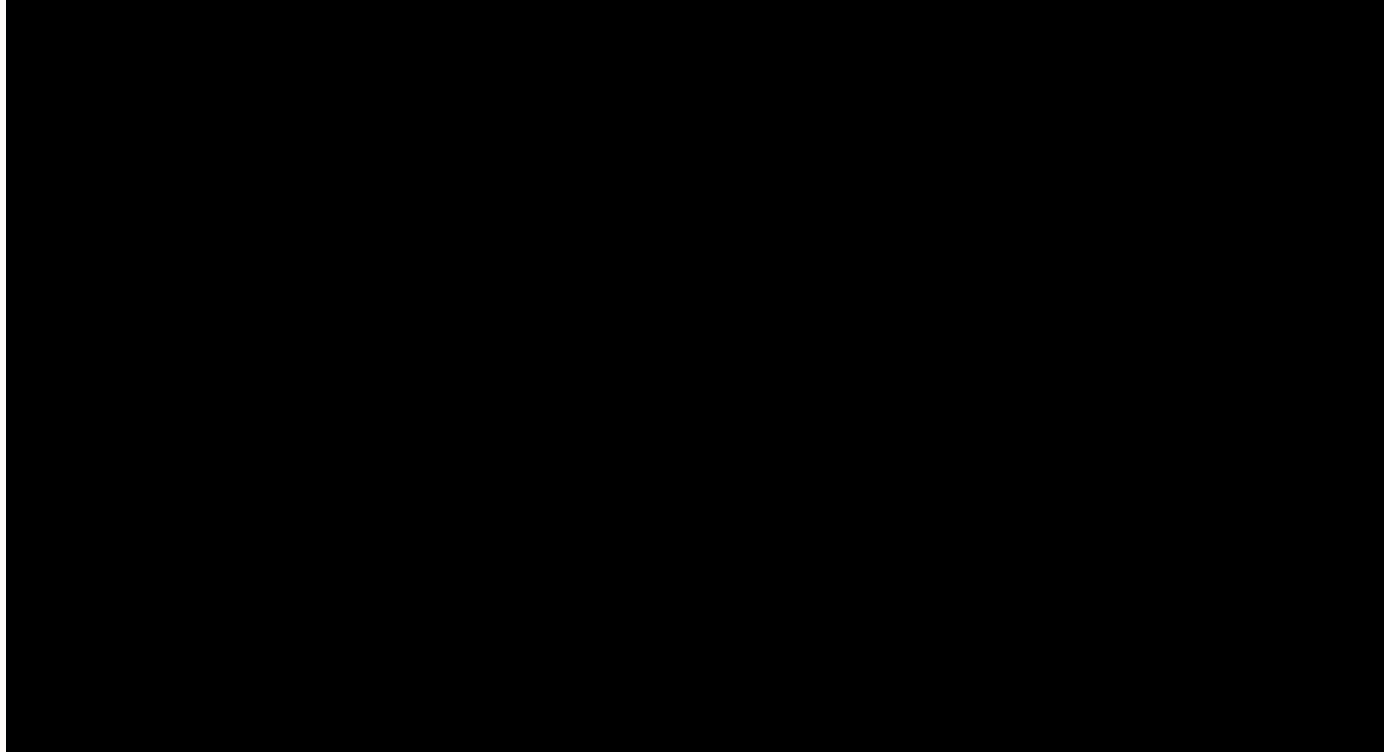
Chatbot Minimum Viable Product (MVP) Demo: Answering Real Business Questions in Seconds



"What is the percentage of female respondents?"

"How do student rate the price range?"

"What is the avg. rating for the quality of Starbucks for different customer groups?"



RCH Empowers Clients, Accelerates Analyst Productivity, and Drives Sales Innovation

Client Self Service Portal (Reducing Cost)

-> **directly access and explore real-time survey data** without needing analyst support.

-> **empowers non-technical users**, shortens decision cycles, and enhances client satisfaction

For example, a brand manager can ask, "How did Gen Z respond to our latest campaign?" and instantly receive summaries, charts, and insights.

Analyst Productivity Booster (Quick Turn-around Time)

Internal research teams often **spend hours manually pulling data and creating reports**.

-> **converts natural language into SQL queries and visuals**, handling standard requests efficiently.

-> **allows analysts to focus on higher-value tasks**, reducing turnaround times and improving overall team productivity.

Sales Enablement and Differentiation (Value-added)

In client pitches and RFPs, the chatbot serves as a powerful demonstration of HarrisX's innovation.

-> showcases the ability to deliver fast, AI-driven insights, **helping the firm stand out from competitors**.

-> **strengthens HarrisX's positioning** as a tech-forward partner in the market research industry.

Maximizing ROI: RCH Delivers Cost Savings and Scalable Value Across the Business

Metric	Before	After	Savings	Assumption
Avg queries per month	500	500	-	Based on team interviews
Time per query (min)	30	5	25	Based on team interviews and prototype test
Time cost per month (hr)	250	42	208	= Time per query * 500 queries per month
Analyst hourly cost (\$/hr)	60	60	-	Estimated fully loaded cost of a mid-level analyst
Monthly cost (\$)	\$15,000	\$2,500	\$12,500	= Time cost per month * Analyst hourly cost
Yearly cost (\$)	\$180,000	\$30,000	\$150,000	= Monthly cost * 12 months

\$100,000

Investment Costs

2,500 hrs

Avg. Time Savings Per Year

\$150,000

Avg. Cost Savings Per Year

50 %

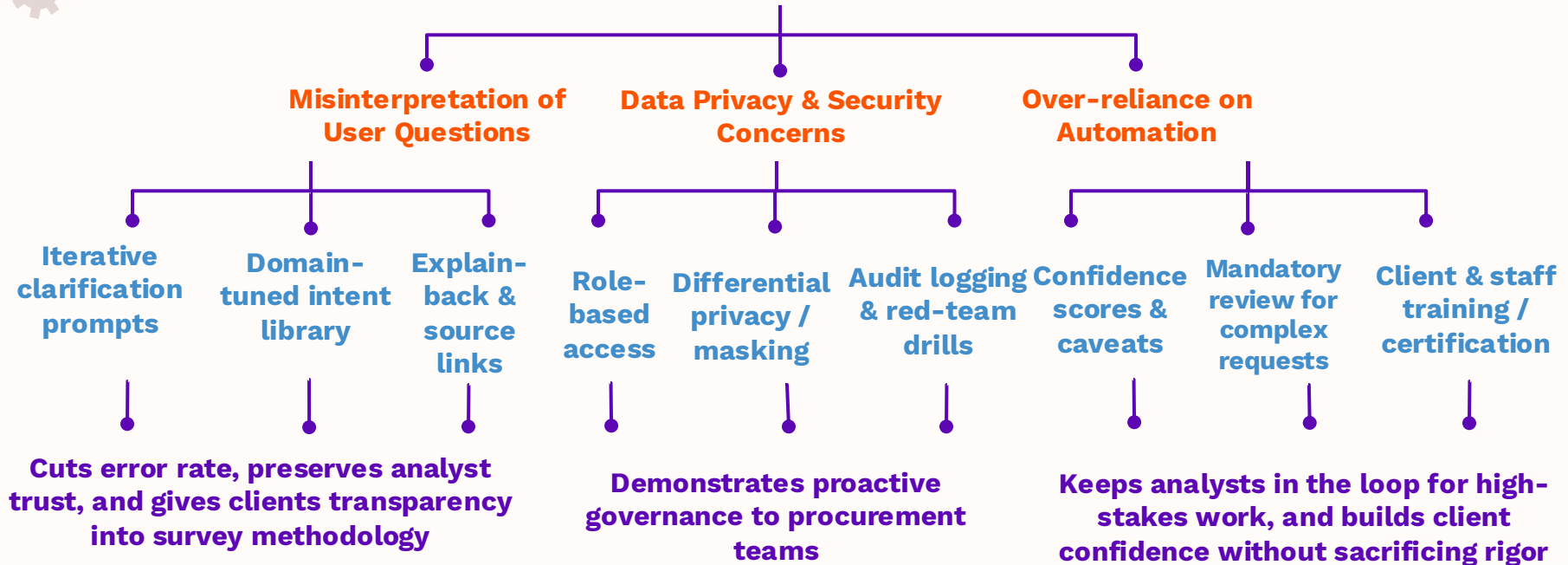
Return on Investments (Year 1 ROI)

8 months

Break-Even Point

From Risks to Resilience: Proactive Mitigations and Impact Analysis for RCH

Risks & Mitigations & Impacts



Appendix: Dataset Snapshot

```
[10] #View first few rows of data
df.head()
```

	Timestamp	1. Your Gender	2. Your Age	3. Are you currently....?	4. What is your annual income?	5. How often do you visit Starbucks?	6. How do you usually enjoy Starbucks?	7. How much time do you normally spend during your visit?	8. The nearest Starbucks outlet to you is...?	9. Do you have Starbucks membership card?	10. What do you most frequently purchase at Starbucks?	11. On average, how much would you spend at Starbucks per visit?	12. How would you rate the quality of Starbucks compared to other brands (Coffee Bean, Old Town White Coffee..) to be:	13. How would you rate the price range at Starbucks?	14. How important are sales and promotions in your purchase decision?	15. How would you rate the ambiance at Starbucks? (lighting, music, etc...)	16. You rate the WiFi quality at Starbucks as..	17. How would you rate the service at Starbucks? (Promptness, friendliness, etc..)	18. How likely you will choose Starbucks for doing business meetings or hangout with friends?	19. How do you come to hear of promotions at Starbucks? Check all that apply.	20. Will you continue buying at Starbucks?
0	2019/10/01 12:38:43 PM GMT+8	Female	From 20 to 29	Student	Less than RM25,000	Rarely	Dine in	Between 30 minutes to 1 hour	within 1km	Yes	...	Less than RM20	4	3	5	5	4	4	3	Starbucks Website/Apps/Social Media,Emails,Dea...	Yes
1	2019/10/01 12:38:54 PM GMT+8	Female	From 20 to 29	Student	Less than RM25,000	Rarely	Take away	Below 30 minutes	1km - 3km	Yes	...	Less than RM20	4	3	4	4	4	5	2	Social Media/in Store displays	Yes
2	2019/10/01 12:38:56 PM GMT+8	Male	From 20 to 29	Employed	Less than RM25,000	Monthly	Dine in	Between 30 minutes to 1 hour	more than 3km	Yes	...	Less than RM20	4	3	4	4	4	4	3	In Store displays,Billboards	Yes
3	2019/10/01 12:39:08 PM GMT+8	Female	From 20 to 29	Student	Less than RM25,000	Rarely	Take away	Below 30 minutes	more than 3km	No	...	Less than RM20	2	1	4	3	3	3	3	Through friends and word of mouth	No
4	2019/10/01 12:39:20 PM GMT+8	Male	From 20 to 29	Student	Less than RM25,000	Monthly	Take away	Between 30 minutes to 1 hour	1km - 3km	No	...	Around RM20 - RM40	3	3	4	2	2	3	3	Starbucks Website/Apps/Social Media	Yes

```
[ ] # Check if any column has missing values
print(df.isnull().sum())
```

```
Timestamp 0
1. Your Gender 0
2. Your Age 0
3. Are you currently....? 0
4. What is your annual income? 0
5. How often do you visit Starbucks? 0
6. How do you usually enjoy Starbucks? 1
7. How much time do you normally spend during your visit? 0
8. The nearest Starbucks's outlet to you is...? 0
9. Do you have Starbucks membership card? 0
10. What do you most frequently purchase at Starbucks? 0
11. On average, how much would you spend at Starbucks per visit? 0
12. How would you rate the quality of Starbucks compared to other brands (Coffee Bean, Old Town White Coffee..) to be: 0
13. How would you rate the price range at Starbucks? 0
14. How important are sales and promotions in your purchase decision? 0
15. How would you rate the ambiance at Starbucks? (lighting, music, etc...) 0
16. You rate the WiFi quality at Starbucks as.. 0
17. How would you rate the service at Starbucks? (Promptness, friendliness, etc..) 0
18. How likely you will choose Starbucks for doing business meetings or hangout with friends? 0
19. How do you come to hear of promotions at Starbucks? Check all that apply. 1
20. Will you continue buying at Starbucks? 0
dtype: int64
```

Appendix: End-to-End Pipeline

Table of contents

AIM - Research Chat Bot (RCH)

- Set-Up
 - Check GPU and Memory
 - Access Permission and Import Data
- Prepare the Data
 - Structure data into Pandas DataFrames
 - Clean Data
 - Structure data into SQL databases
 - Sample SQL query
- Prompt Engineer
 - Load the openai API key
- Prompt to SQL
 - SQL to Summary
 - Visualization
- Full Code

[+ Section](#)

(1) Prompt to SQL

```

import re

# Step 1: User input
user_question = input("Welcome! What would you like to know today? ")

# Step 2: Generate SQL using GPT
def make_sql_prompt(user_question, table_name, columns):
    col_str = ', '.join([f'"{col}"' for col in columns])
    prompt = (
        f'You are a SQL expert. '
        f'Here is the schema for the table "{table_name}":\n'
        f'Columns: {col_str}\n'
        f'Based on this schema, write a SQL query to answer the following question:\n'
        f'{user_question}\n'
        "Return only the SQL code, without any markdown formatting or explanation."
    )
    return prompt

columns = list(df_01.columns)
sql_prompt = make_sql_prompt(user_question, "customer_survey", columns) #further improvements on automatically select the database
sql_query_response = get_completion(sql_prompt, model="gpt-4o")

# Step 3: execute the SQL query and print the results
print("\nSQL Query:\n", sql_query_response)

try:
    result_df = pd.read_sql_query(sql_query_response, conn)
    print("\nSQL Query Result:\n", result_df)
except Exception as e:
    print("Error running SQL:", e)

```

... Welcome! What would you like to know today?

(2) SQL to Summary

```

def make_summary_prompt(user_question, result_df):
    # Convert result_df to a string for context (truncate if too long)
    result_str = result_df.to_string(index=False)
    prompt = (
        f'You are a data analyst. '
        f'The user asked: {user_question}\n'
        f'Here is the query result:\n{result_str}\n'
        "Summarize this result and the key insight in plain English for a business audience."
    )
    return prompt

summary_prompt = make_summary_prompt(user_question, result_df)
summary = get_completion(summary_prompt, model="gpt-4o")
print("Summary:\n", summary)

```

Summary:
The survey results indicate that among the respondents, there are 65 females and 56 males. This means that

Chart Recommendation

```

def chart_recom(user_question, result_df):
    # Convert result DataFrame to string (truncate if very large)
    result_str = result_df.head(10).to_string(index=False) # show up to 10 rows
    prompt = (
        f'You are a Business Intelligence Visualization Analyst. '
        f'The user asked: {user_question}\n'
        f'Here are the query results:\n{result_str}\n'
        "Recommend the most appropriate chart or visualization for these results (e.g., seaborn/matplotlib chart)"
    )
    return prompt

summary_prompt = chart_recom(user_question, result_df)
summary_and_chart = get_completion(summary_prompt, model="gpt-4o")

def extract_chart_type(gpt_response):
    chart_types = [
        'bar', 'line', 'scatter', 'histogram', 'box', 'box & whisker', 'violin',
        'treemap', 'heatmap', 'swarm', 'strip', 'bullet', 'packed bubble'
    ]
    for chart in chart_types:
        if re.search(rf'\b(re.escape(chart))\b', gpt_response.lower()):
            return chart
    return None

# Main visualization function
def plot_flexible_chart(result_df, chart_type):
    plt.close('all')
    cols = result_df.columns.tolist()
    chart_type_lower = chart_type.lower() if chart_type else None

    # Map common chart names to seaborn/matplotlib functions
    chart_map = {
        "bar": lambda: sns.barplot(x=cols[0], y=cols[1], data=result_df, hue=cols[0], palette="Purples", legend=False),
        "line": lambda: sns.lineplot(x=cols[0], y=cols[1], data=result_df, color=sns.color_palette("Purples")[4]),
        "scatter": lambda: sns.scatterplot(x=cols[0], y=cols[1], data=result_df, color=sns.color_palette("Purples")[5]),
        "histogram": lambda: sns.histplot(result_df[cols[0]], color=sns.color_palette("Purples")[3]),
        "box": lambda: sns.boxplot(y=result_df[cols[0]], color=sns.color_palette("Purples")[6]),
        "box & whisker": lambda: sns.boxplot(y=result_df[cols[0]], color=sns.color_palette("Purples")[6]),
        "violin": lambda: sns.violinplot(y=result_df[cols[0]], color=sns.color_palette("Purples")[5]),
        "treemap": lambda: squarify.plot(
            sizes=result_df[cols[1]],
            label=result_df[cols[0]].astype(str),
            color=sns.color_palette("Purples", len(result_df))),
        "heatmap": lambda: sns.heatmap(result_df.corr(), annot=True, cmap="Purples"),
        "swarm": lambda: sns.swarmplot(x=cols[0], y=cols[1], data=result_df, color=sns.color_palette("Purples")[5]),
        "strip": lambda: sns.stripplot(x=cols[0], y=cols[1], data=result_df, color=sns.color_palette("Purples")[6])
    }

    # Try to plot if mapping exists and data shape fits
    try:
        if chart_type_lower in chart_map:
            chart_map[chart_type_lower]()
            plt.tight_layout()
            plt.show()
        else:
            print("No visuals recommended.")
    except Exception:
        print("No visuals recommended.")

chart_type = extract_chart_type(summary_and_chart)
if chart_type:
    plot_flexible_chart(result_df, chart_type)
else:
    print("No visuals recommended.")

```

Appendix: RCH Prototype Demo



🔥 Welcome to HarrisX Research Chatbot!

I'm here to help you explore customer survey data and generate insights.

Type your business question below to get started. For example:

- What is the average satisfaction rating by gender?
- How do students rate the price range?
- What is average rating of the quality of Starbucks compared to other brands for different customer groups?

Enter your question here: How do students rate the price range?

Generated SQL Query

```
SELECT
    "13. How would you rate the price range at Starbucks?" AS price_range_rating,
    COUNT(*) AS student_count
FROM
    customer_survey
WHERE
    "3. Are you currently...?" = 'Student'
GROUP BY
    "13. How would you rate the price range at Starbucks?"
ORDER BY
    price_range_rating;
```

SQL Query Result (top 10 rows):

price_range_rating	student_count	
0	1	5
1	2	10
2	3	18
3	4	6
4	5	3

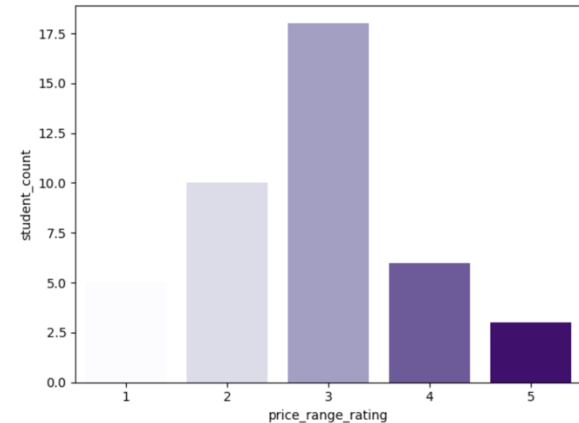
Summary & Key Insight

The data shows how students rate the price range on a scale from 1 to 5, with 1 being the lowest and 5 being the highest. Here's a summary of the findings:

- A small number of students (5) rated the price range as 1, indicating they found it very affordable.
- More students (10) rated it as 2, suggesting it is still relatively affordable but slightly more expensive than the lowest rating.
- The majority of students (18) rated the price range as 3, indicating a moderate level of affordability.
- Fewer students (6) rated it as 4, suggesting it is on the higher end of the price spectrum.
- Only a small number of students (3) rated it as 5, indicating they found it to be the most expensive.

Key Insight: Most students perceive the price range as moderately affordable, with the highest concentration of ratings at level 3. This suggests that while the pricing is generally acceptable, there is a significant portion of students who find it more expensive, as indicated by the ratings of 4 and 5. Businesses might consider this feedback when evaluating their pricing strategy to ensure it aligns with student expectations and budgets.

Visualization: Bar Chart



Appendix: Financial Benefit & ROI Calculation

Metric	Before	After	Savings	Assumption
Avg queries per month	500	500	-	Based on team interviews
Time per query (min)	30	5	25	Based on team interviews and prototype test
Time cost per month (hr)	250	42	208	= Time per query * 500 queries per month
Analyst hourly cost (\$/hr)	60	60	-	Estimated fully loaded cost of a mid-level analyst
Monthly cost (\$)	\$15,000	\$2,500	\$12,500	= Time cost per month * Analyst hourly cost
Yearly cost (\$)	\$180,000	\$30,000	\$150,000	= Monthly cost * 12 months

Investment	\$100,000
Yearly Time Savings	2,500
Yearly Cost Savings	\$150,000
Return on Investment	50.00%
Break-even point (months)	8