UNITED

UNITED AIRLINES HACKATHON-SKYHACK 2.0

TEAMMATES

1.MOHD SHOHEL
2.DEEPAK KUMAR

Department of computer science,University Of Delhi

STEPS

Load Dataset → **Preprocessing** → **Feature Eengineering** → **DATA VISULAIZATION** → **DELIVERABLES**

Load Dataset

Load all csv files and understand the attribute present in these files

Preprocessing

Using regex removing special characters,trailing and leading spaces from primary_call_reason in reason dataset.Also finding missing and NAN values in customer dataset elite_level_code and fill with mode value and similarly done for sentiment dataset

Feature Eengineering

calculation of

AHT = Total Handle Time / Total Number of Calls

AST = Total Waiting Time / Total Number of Calls

with the help of call_start_datetime, agent_assigned_datetime and call_end_datetime

DATA VISULAIZATION

Using Barplot ,Bargraph and Boxplot

DELIVERABLES

DELIVERABLE 1

Explore the factors contributing to extended call durations, such as agent performance, call types, and sentiment. Identify key drivers of long AHT and AST, especially during high volume call periods.

Agent Performance-: The tone and behavior of agents during calls can significantly influence call durations. Agents who maintain a calm and friendly tone may handle calls more efficiently, whereas those who are neutral or frustrated may lead to longer calls.

Call types:-Different types of calls, categorized by the primary reason for the call (i.e checkouts,irrops), may require varying levels of complexity. Calls that involve more detailed explanations or troubleshooting are likely to take longer.The most frequent call reason is irrops and least frequent call reason is unaccompanied minor.

Customer Sentiment:-The sentiment of customers (i.e frustrated, calm) can greatly influence both AHT and AST. Frustrated customers may require more time and attention, resulting in longer AHT. Additionally, higher silence percentages during calls may indicate hesitation or confusion from the customer.

UNITED

Total Handle Time Of All Calls (sec): 46468920.0
Total Waiting Time Of All Calls (sec): 29125500.0
Total Number Of Calls: 66653

The Average Handle Time Of Calls (sec) 697.1767212278517
The Average Speed To Answer Time Of Calls(sec) 436.97207927625163

Calculating Average Handle Time And Average Speed Time To Answer

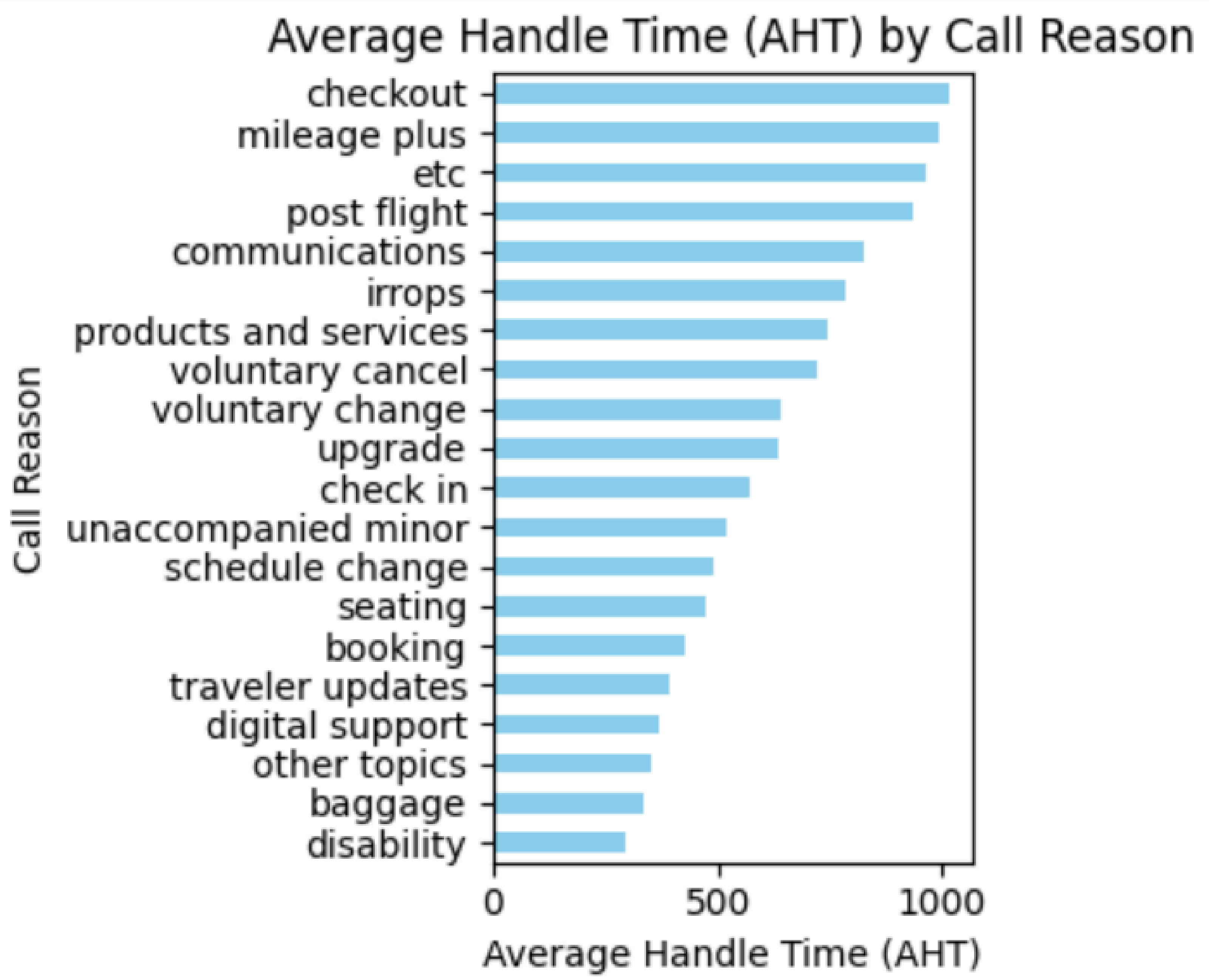
```
[ ] # Calculate AHT and AST
    total_handle_time = merged_data['Handle_Time'].sum()
    total_waiting_time = merged_data['Waiting_Time'].sum()
    total_calls = merged_data.call_id.value_counts().sum()
    print("Total Handle Time Of All Calls (sec):",total_handle_time)
    print("Total Waiting Time Of All Calls (sec):",total_waiting_time)
    print("Total Number Of Calls:",total_calls)
```

```
Total Handle Time Of All Calls (sec): 46468920.0
Total Waiting Time Of All Calls (sec): 29125500.0
Total Number Of Calls: 66653
```

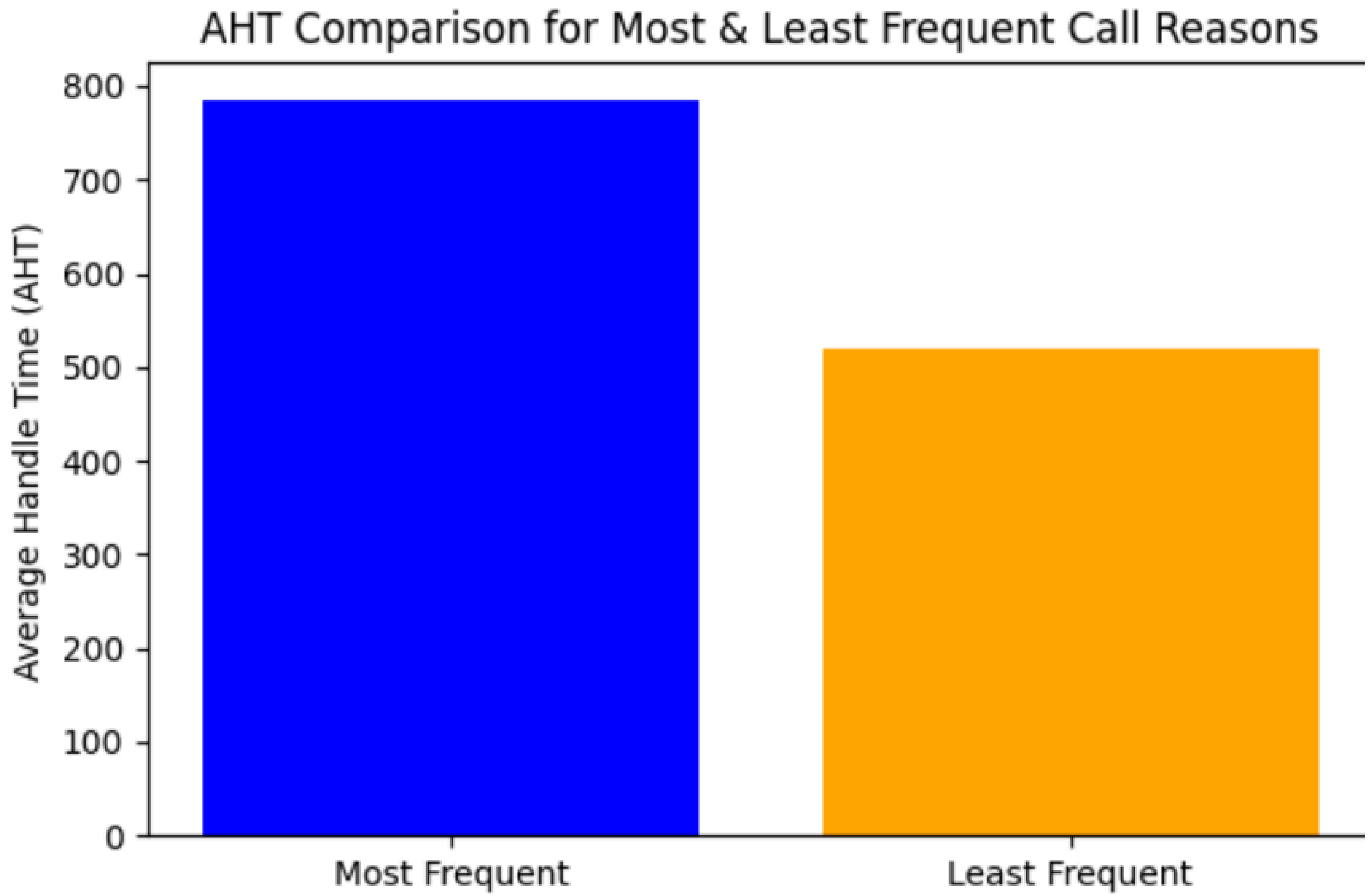
```
AHT = total_handle_time / total_calls
print("The Average Handle Time Of Calls (sec)",AHT)
AST = total_waiting_time / total_calls
print("The Average Speed To Answer Time Of Calls(sec)",AST)
```

```
The Average Handle Time Of Calls (sec) 697.1767212278517
The Average Speed To Answer Time Of Calls(sec) 436.97207927625163
```

INSIGHTS

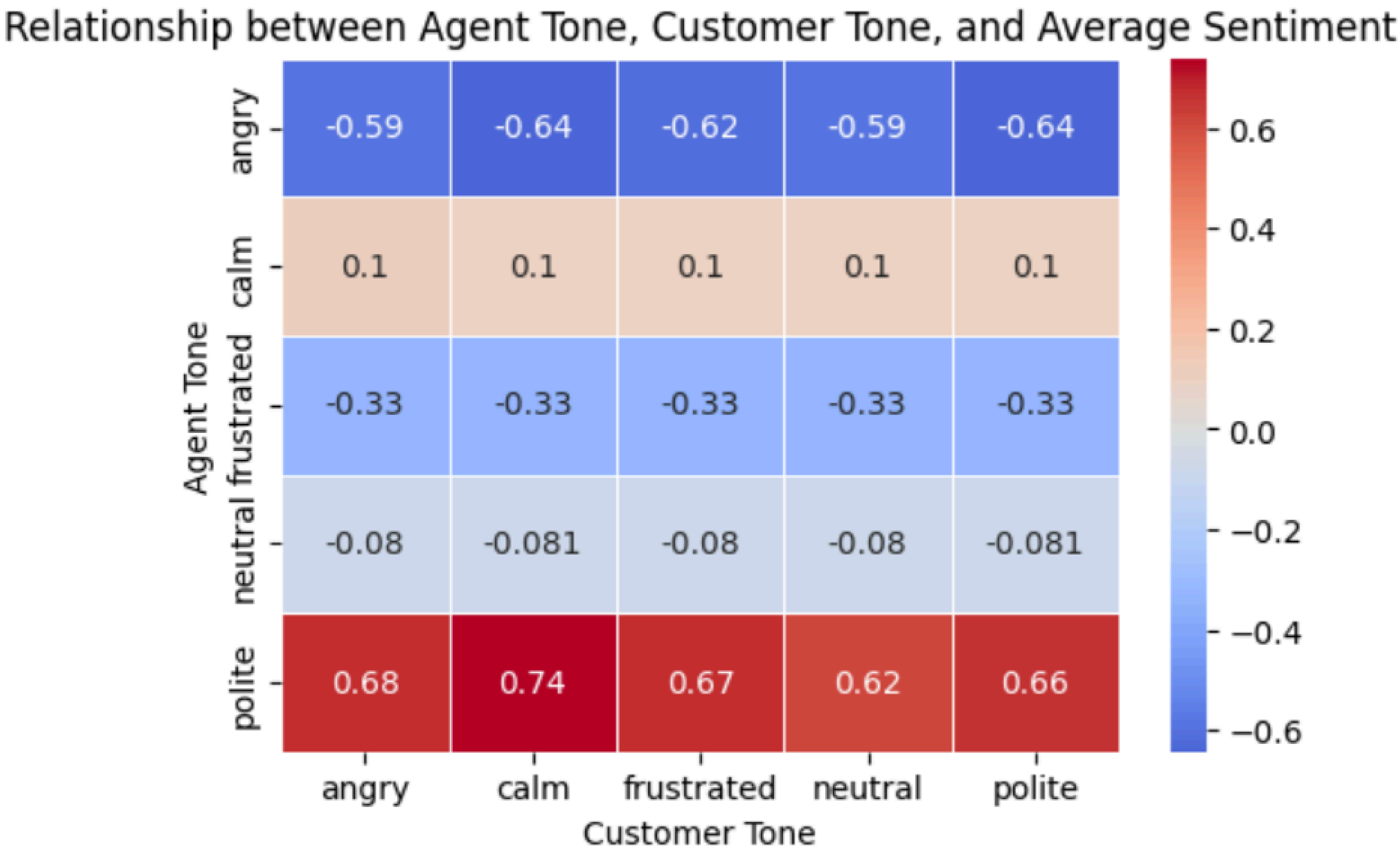


Average Handle Time (AHT) by Call Reason
checkout
mileage plus
etc
post flight
communications
irrops
products and services
voluntary cancel
voluntary change
upgrade
check in
unaccompanied minor
schedule change
seating
booking
traveler updates
digital support
other topics
baggage
disability
Call Reason
0
500
1000
Average Handle Time (AHT)

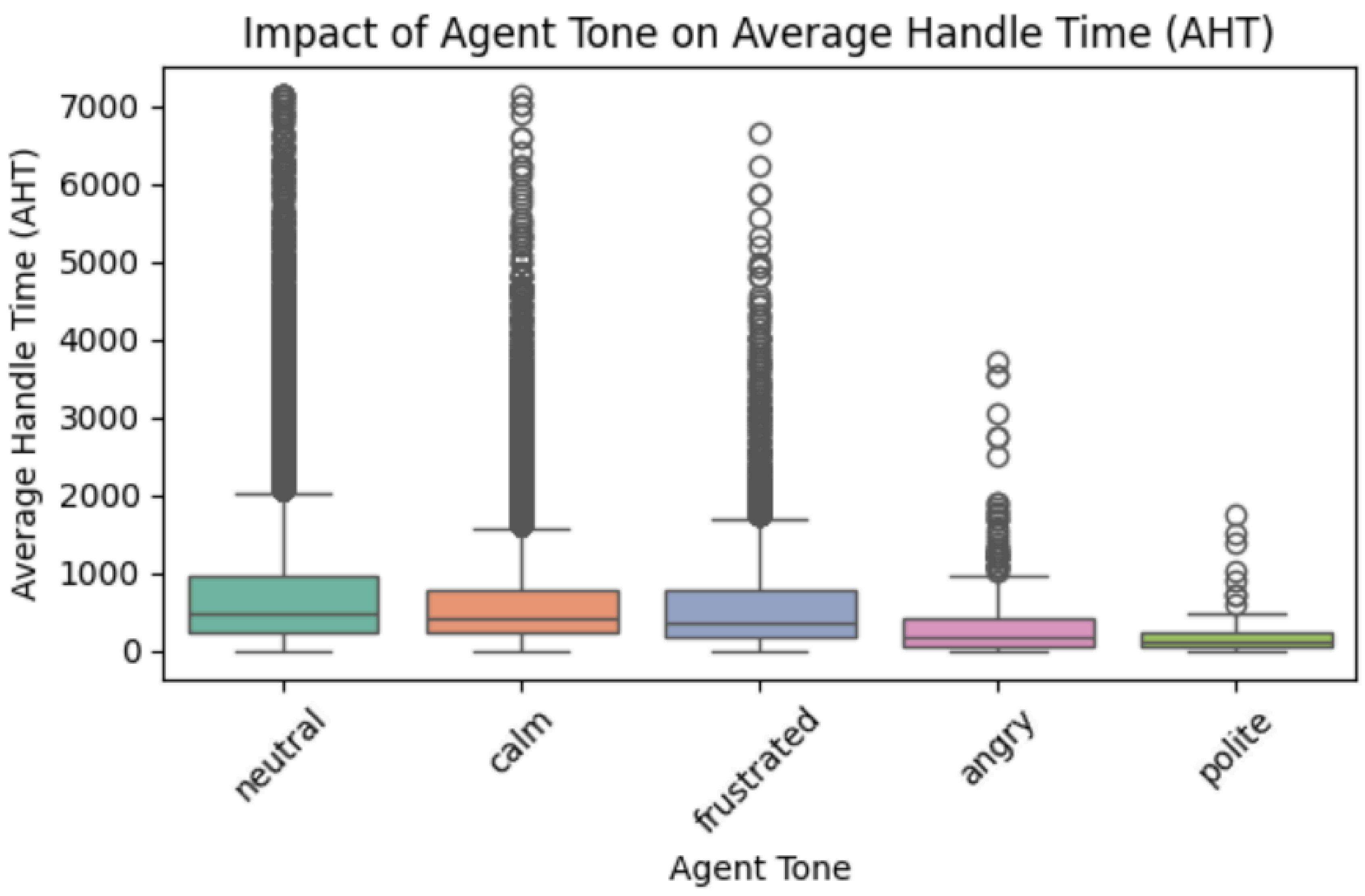


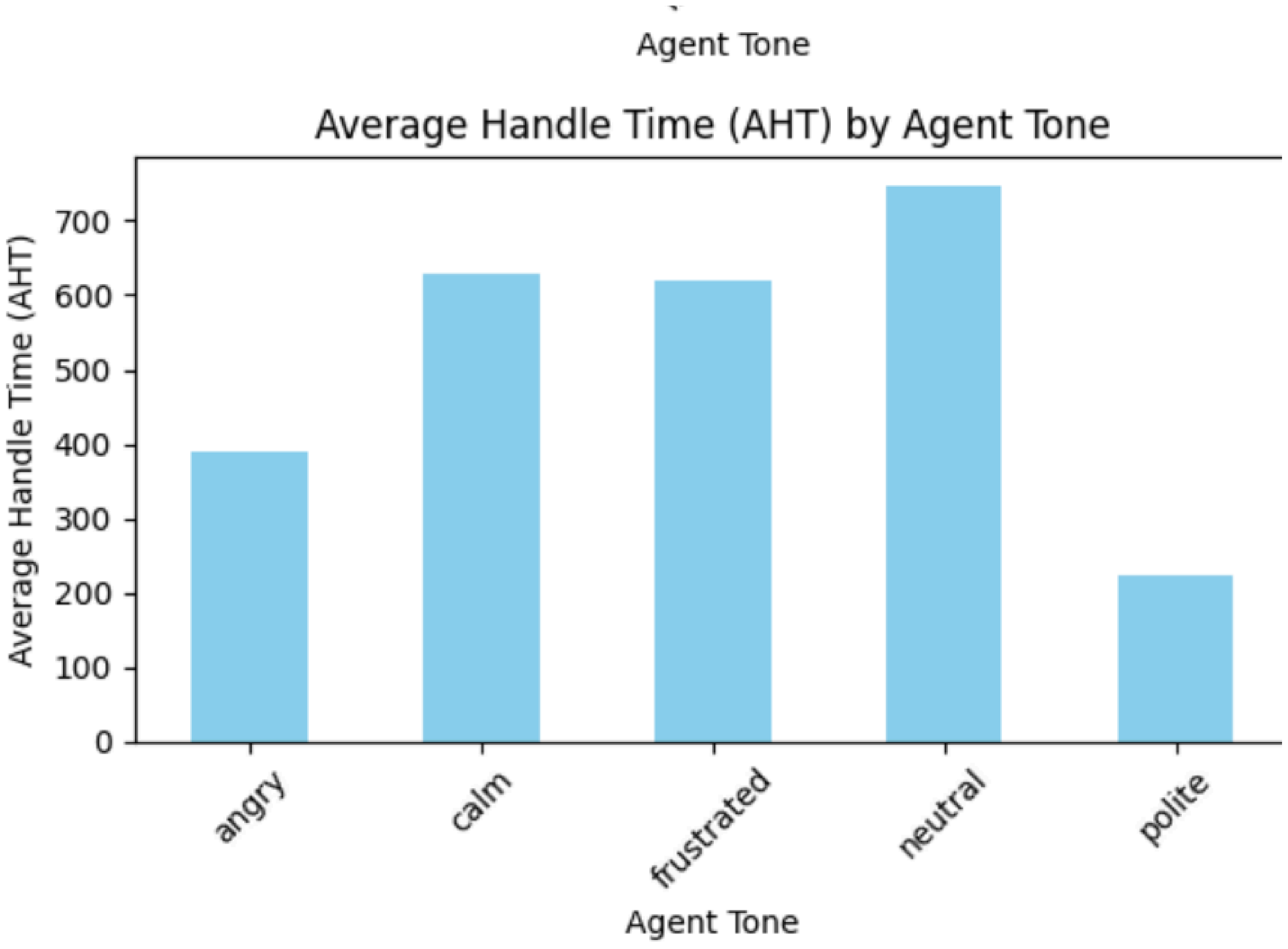
AHT Comparison for Most & Least Frequent Call Reasons
Average Handle Time (AHT)
800
700
600
500
400
300
200
100
0
Most Frequent
Least Frequent

INSIGHTS

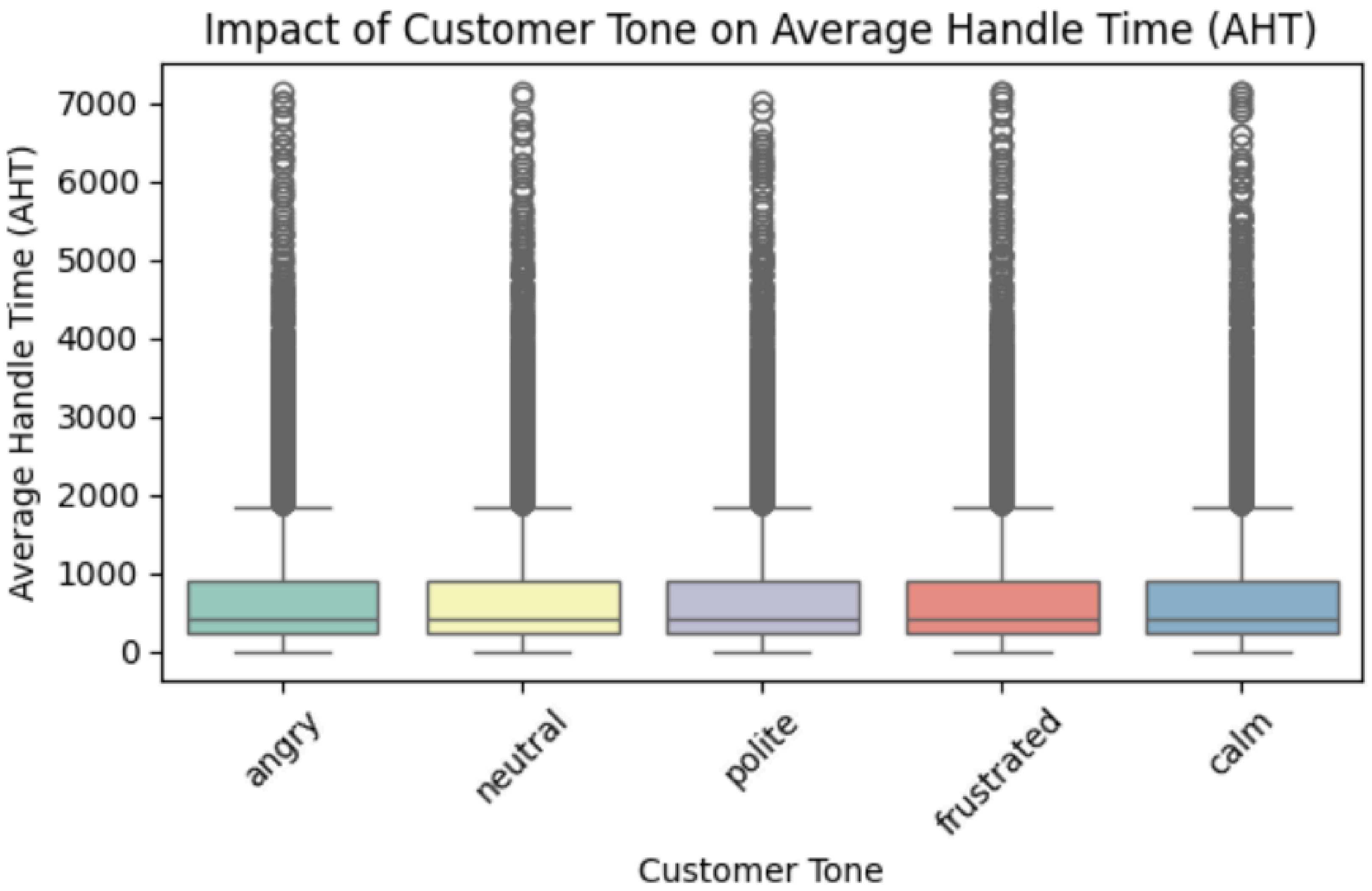

Relationship between Agent Tone, Customer Tone, and Average Sentiment
Agent Tone
angry
calm
frustrated
neutral
polite
Customer Tone
angry
calm
frustrated
neutral
polite
-0.59
-0.64
-0.62
-0.59
-0.64
0.1
0.1
0.1
0.1
0.1
-0.33
-0.33
-0.33
-0.33
-0.33
-0.08
-0.081
-0.08
-0.08
-0.081
0.68
0.74
0.67
0.62
0.66
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

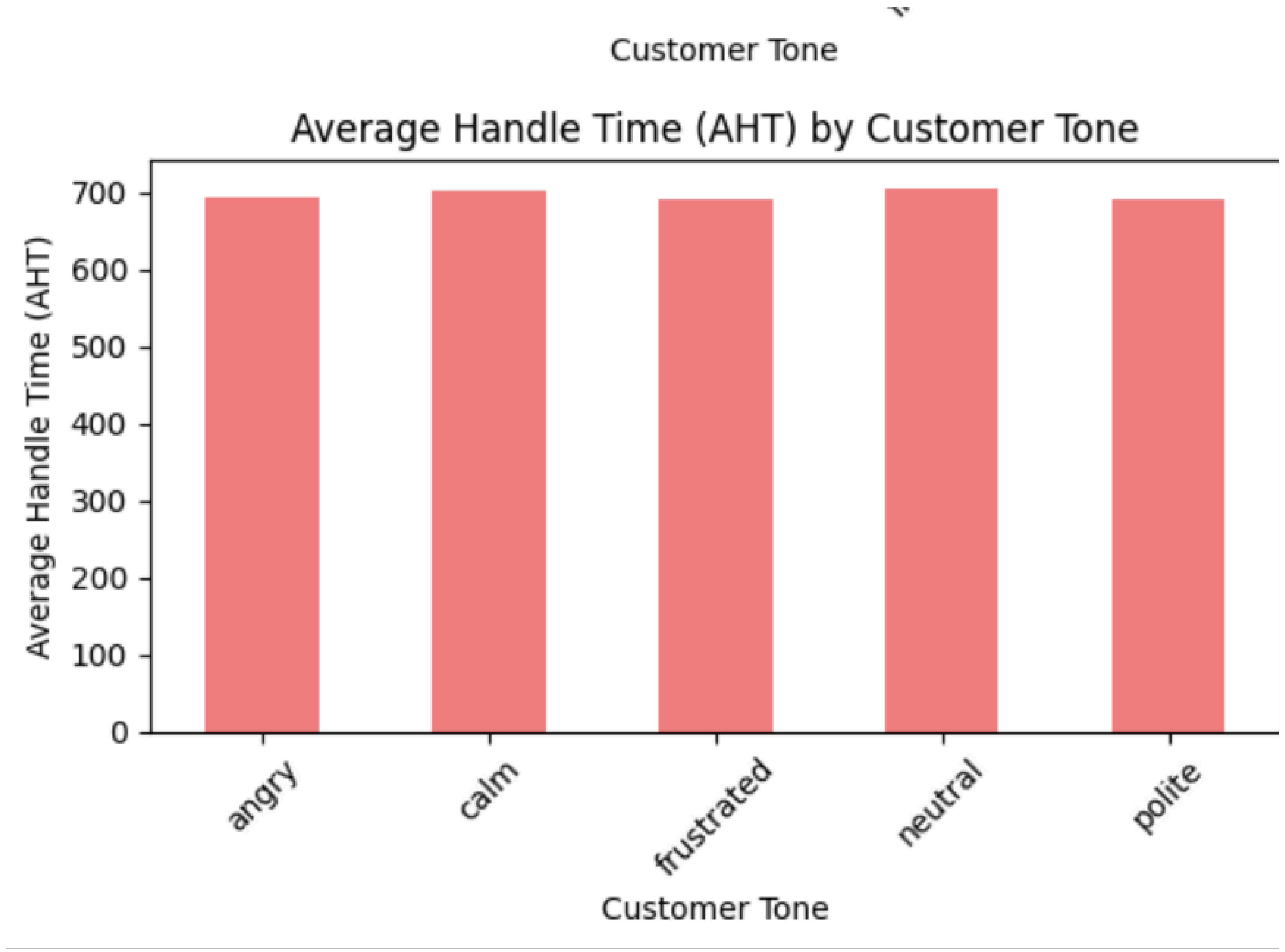
INSIGHTS


Impact of Agent Tone on Average Handle Time (AHT)
Average Handle Time (AHT)
7000
6000
5000
4000
3000
2000
1000
0
neutral
calm
frustrated
angry
polite
Agent Tone

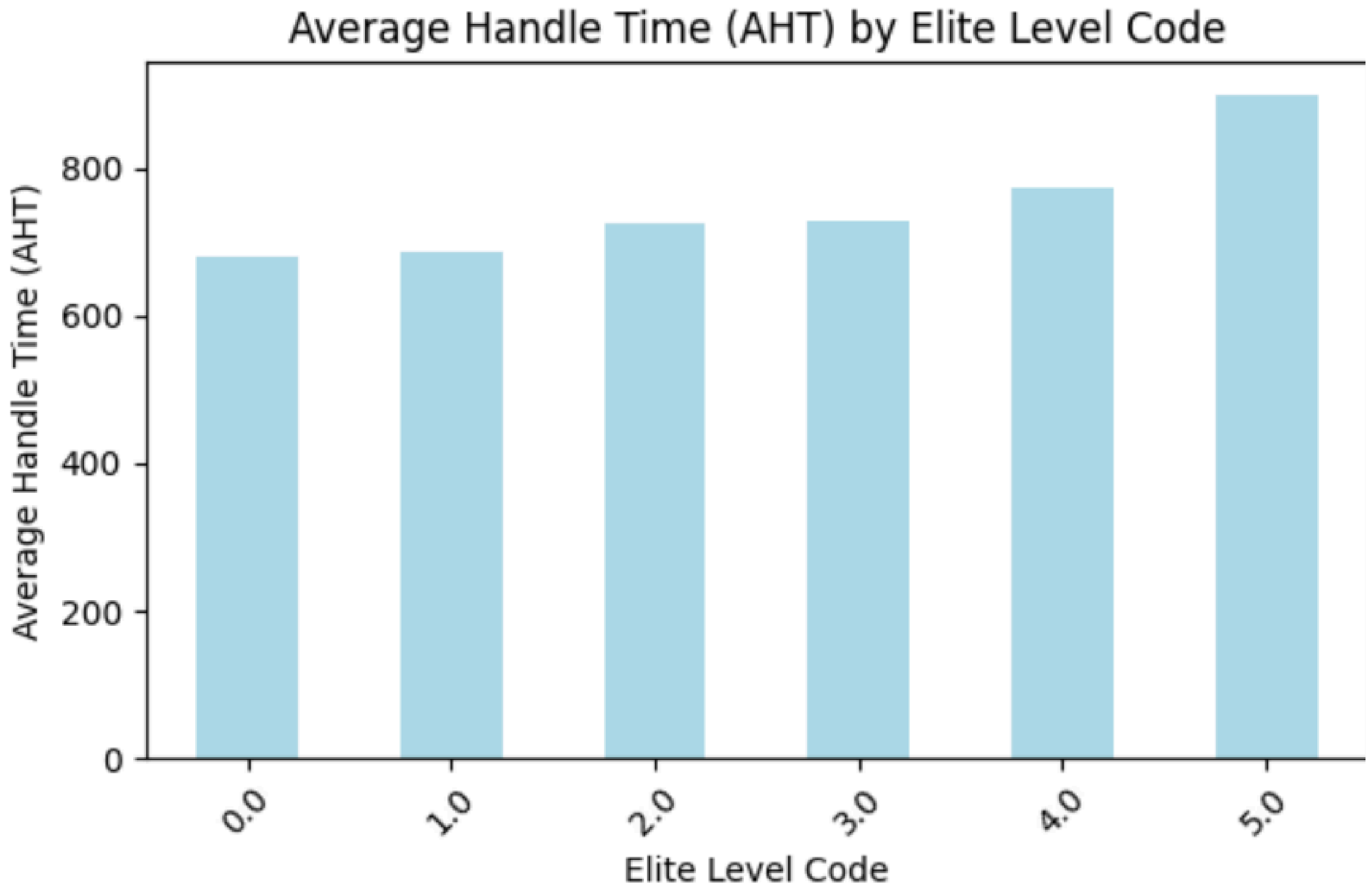

Agent Tone
Average Handle Time (AHT) by Agent Tone
Average Handle Time (AHT)
700
600
500
400
300
200
100
0
angry
calm
frustrated
neutral
polite
Agent Tone

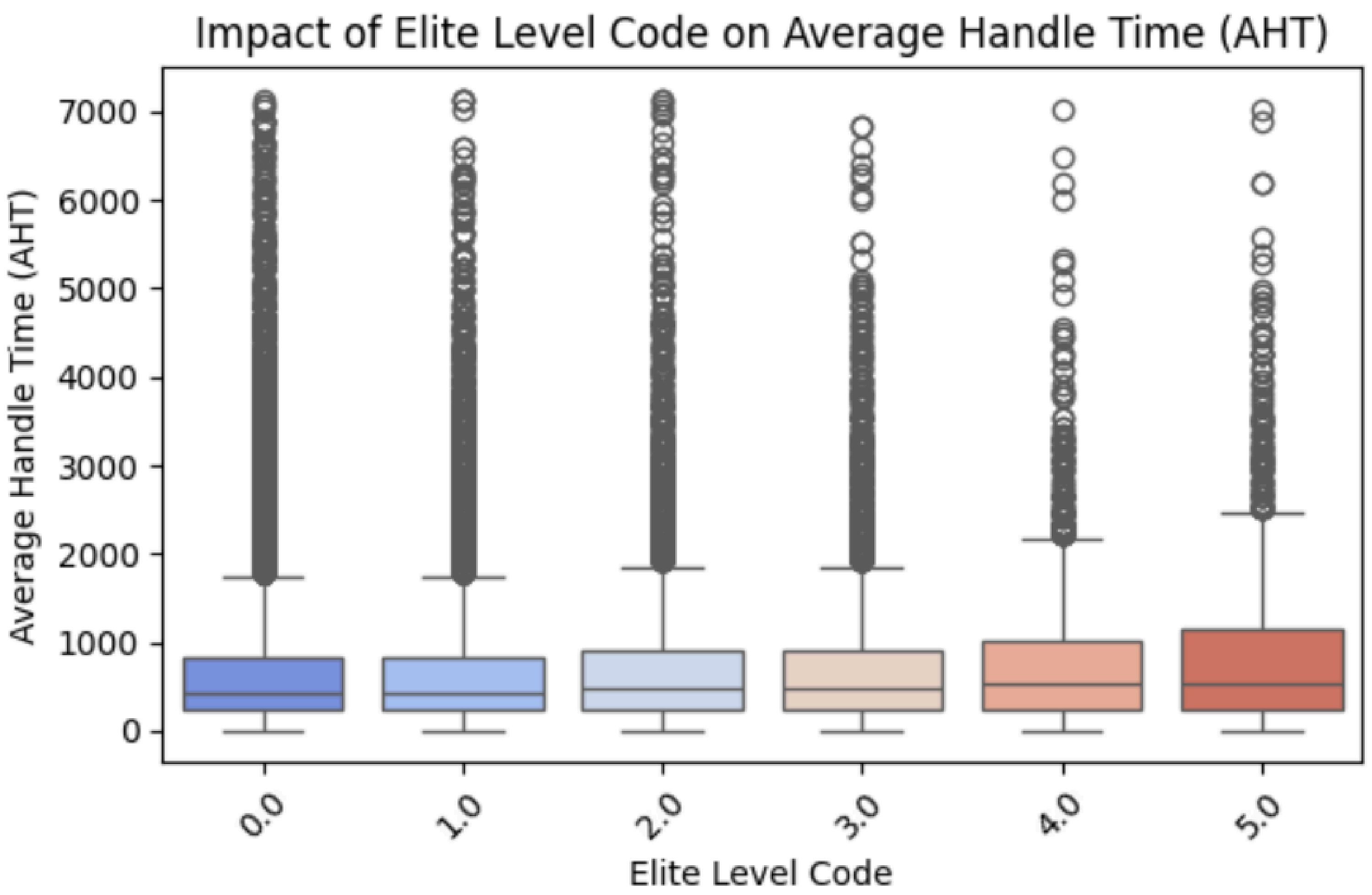
INSIGHTS


Impact of Customer Tone on Average Handle Time (AHT)
Average Handle Time (AHT)
7000
6000
5000
4000
3000
2000
1000
0
angry
neutral
polite
frustrated
calm
Customer Tone


Customer Tone
Average Handle Time (AHT) by Customer Tone
Average Handle Time (AHT)
700
600
500
400
300
200
100
0
angry
calm
frustrated
neutral
polite
Customer Tone

INSIGHTS


Average Handle Time (AHT) by Elite Level Code
Average Handle Time (AHT)
0
200
400
600
800
0.0
1.0
2.0
3.0
4.0
5.0
Elite Level Code


Impact of Elite Level Code on Average Handle Time (AHT)
Average Handle Time (AHT)
0
1000
2000
3000
4000
5000
6000
7000
0.0
1.0
2.0
3.0
4.0
5.0
Elite Level Code

UNITED

Percentage difference between the average handling time for the most frequent and least frequent call reasons

```
Most Frequent Call Reason: irrops, AHT: 785.1160694162722
Least Frequent Call Reason: unaccompanied minor, AHT: 519.2307692307693
Percentage Difference in AHT: 51.21%
```

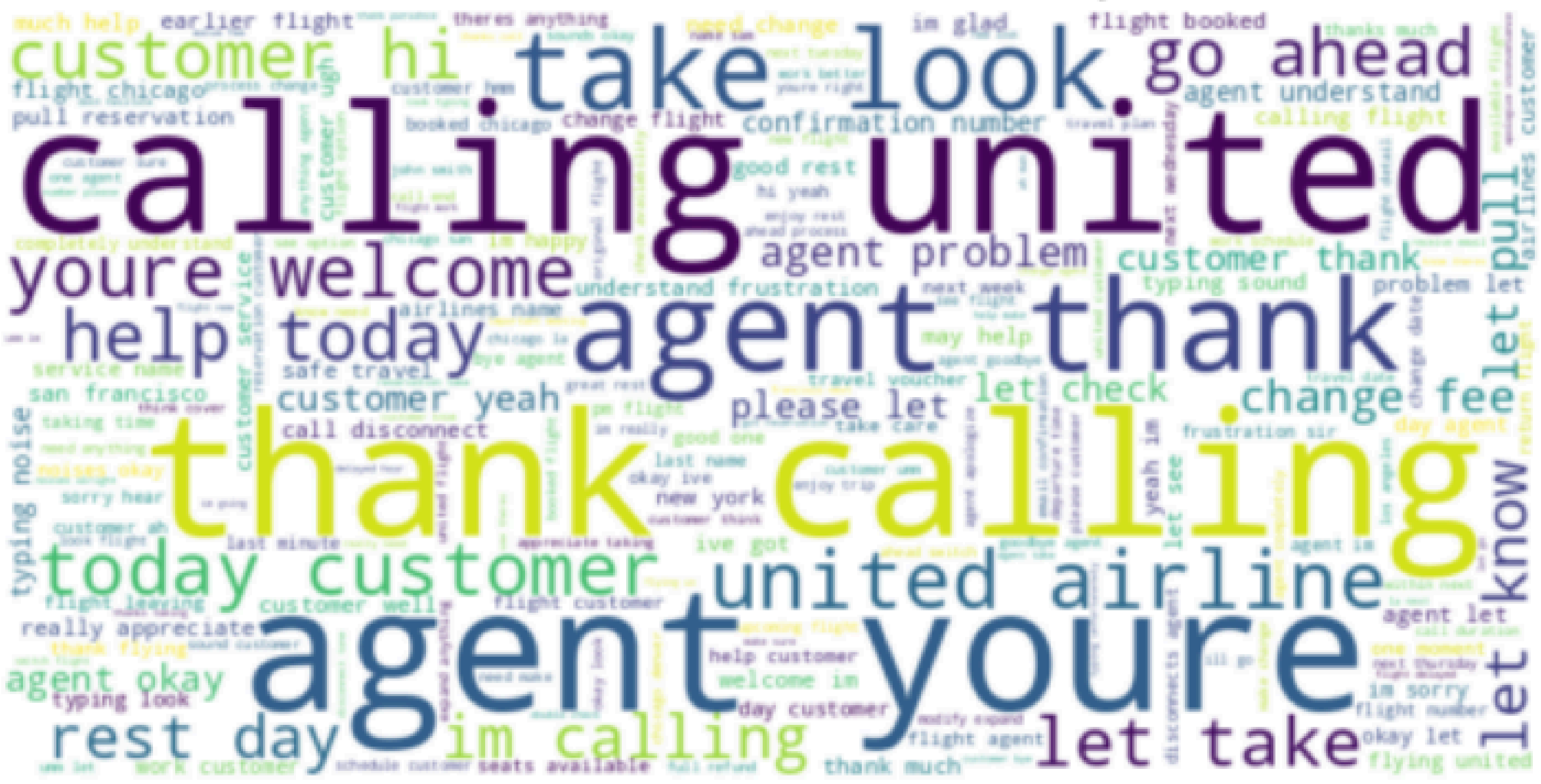
```
Call Reason vs Handle Time:
 primary_call_reason
checkout                  1016.853814
mileage plus               995.573406
etc                        962.899160
post flight                932.896074
communications             826.718750
irrops                     785.116069
products and services      746.560624
voluntary cancel           721.866833
voluntary change           639.153761
upgrade                    632.344777
check in                   574.128151
unaccompanied minor        519.230769
schedule change            490.013680
seating                    474.994501
booking                    427.736064
traveler updates           393.233725
digital support            372.293878
other topics               350.097800
baggage                    333.644068
disability                 292.109181
Name: AHT, dtype: float64
```

DELIVERABLE 2

Analyse the transcripts and call reasons to identify granular reasons associated to recurring problems that could be resolved via self-service options in the IVR system.

We analyze the call trascripts using NLP(natural language precessing) from which removing trailing and leading spaces,special characters and stop words

Word Cloud of Call Transcripts



UNITED

Insights and reason

Top 10 most frequent words in call transcripts:

[('flight', 584640),
('agent', 514677),
('customer', 506450),
('let', 280207),
('change', 227941),
('im', 220868),
('help', 186566),
('like', 180127),
('thank', 171490),
('would', 163331)]

```
Call Reason vs Handle Time:
 primary_call_reason
checkout                     1016.853814
mileage plus                  995.573406
etc                           962.899160
post flight                   932.896074
communications                826.718750
irrops                        785.116069
products and services         746.560624
voluntary cancel              721.866833
voluntary change              639.153761
upgrade                       632.344777
check in                      574.128151
unaccompanied minor           519.230769
schedule change               490.013680
seating                       474.994501
booking                       427.736064
traveler updates              393.233725
digital support               372.293878
other topics                  350.097800
baggage                       333.644068
disability                    292.109181
Name: AHT, dtype: float64
```

Specific improvements to the IVR options to effectively reduce agent intervention in these cases, along with solid reasoning to support your recommendations.

RECOMMENDATIONS

After Analyze the most commonly associated reasons are irrops ,voluntary change ,seating
So we can add these primary call reasons in our IVR system so that mostly customer queries will resolve through IVR systems without agent intervention

```
Recommendations for IVR system improvements:
Consider adding a self-service option for: irrops
Consider adding a self-service option for: voluntary change
Consider adding a self-service option for: seating
```

DELIVERABLE 3

1. Data Preprocessing

- Merged multiple datasets and handled missing values (e.g., filling average_sentiment with the mode,elite_level_code NAN values with mode and agent tone NAN values with mode).
- Created new features like call_duration and vectorized call_transcript using TF-IDF.
- Normalized primary_call_reason and encoded categorical features.

2. EDA

- Visualized distributions of primary_call_reason and explored relationships between sentiment and call reasons.

3. Model Training

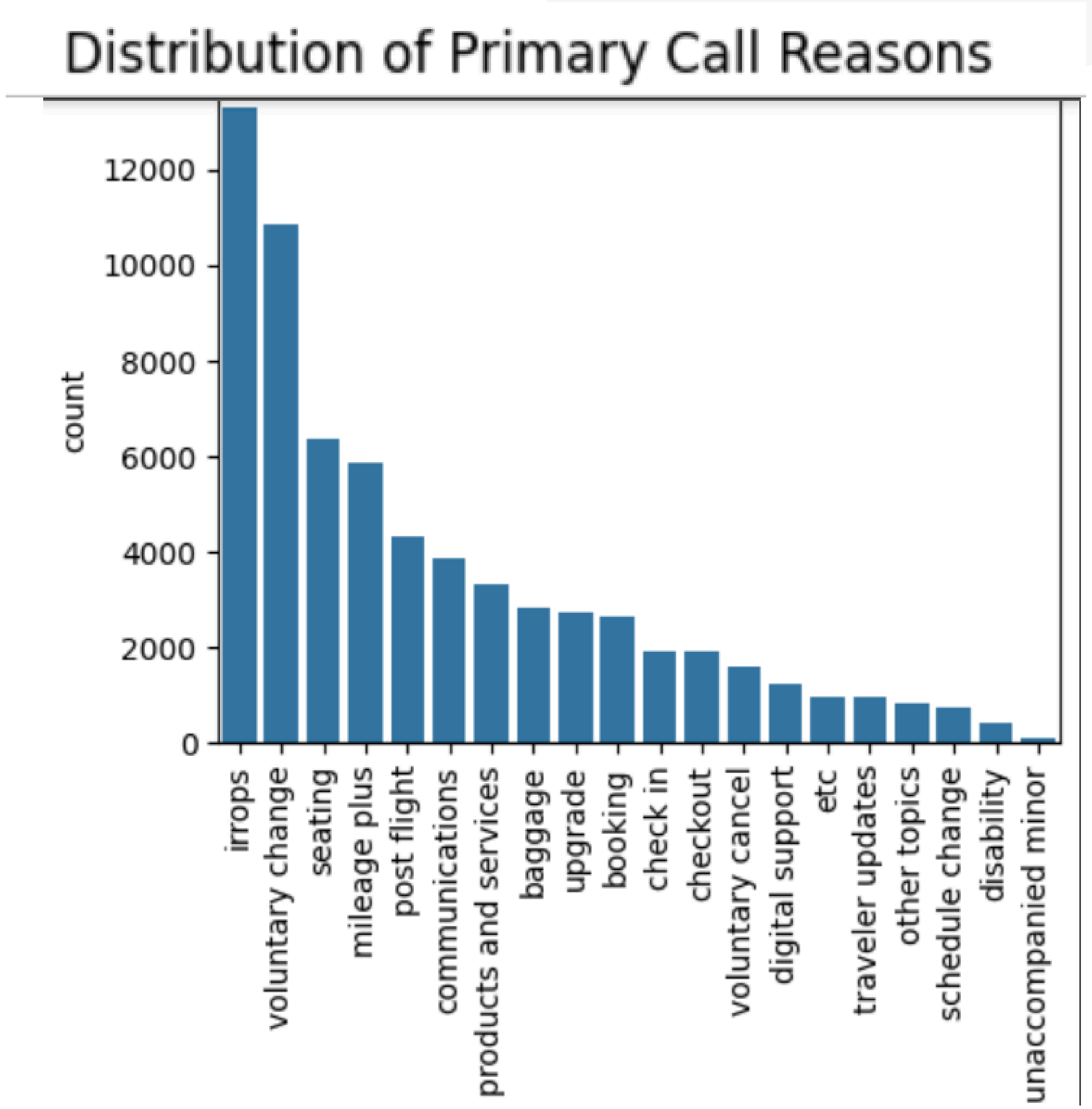
- Trained a Random Forest Classifier on the dataset split into training/testing sets.

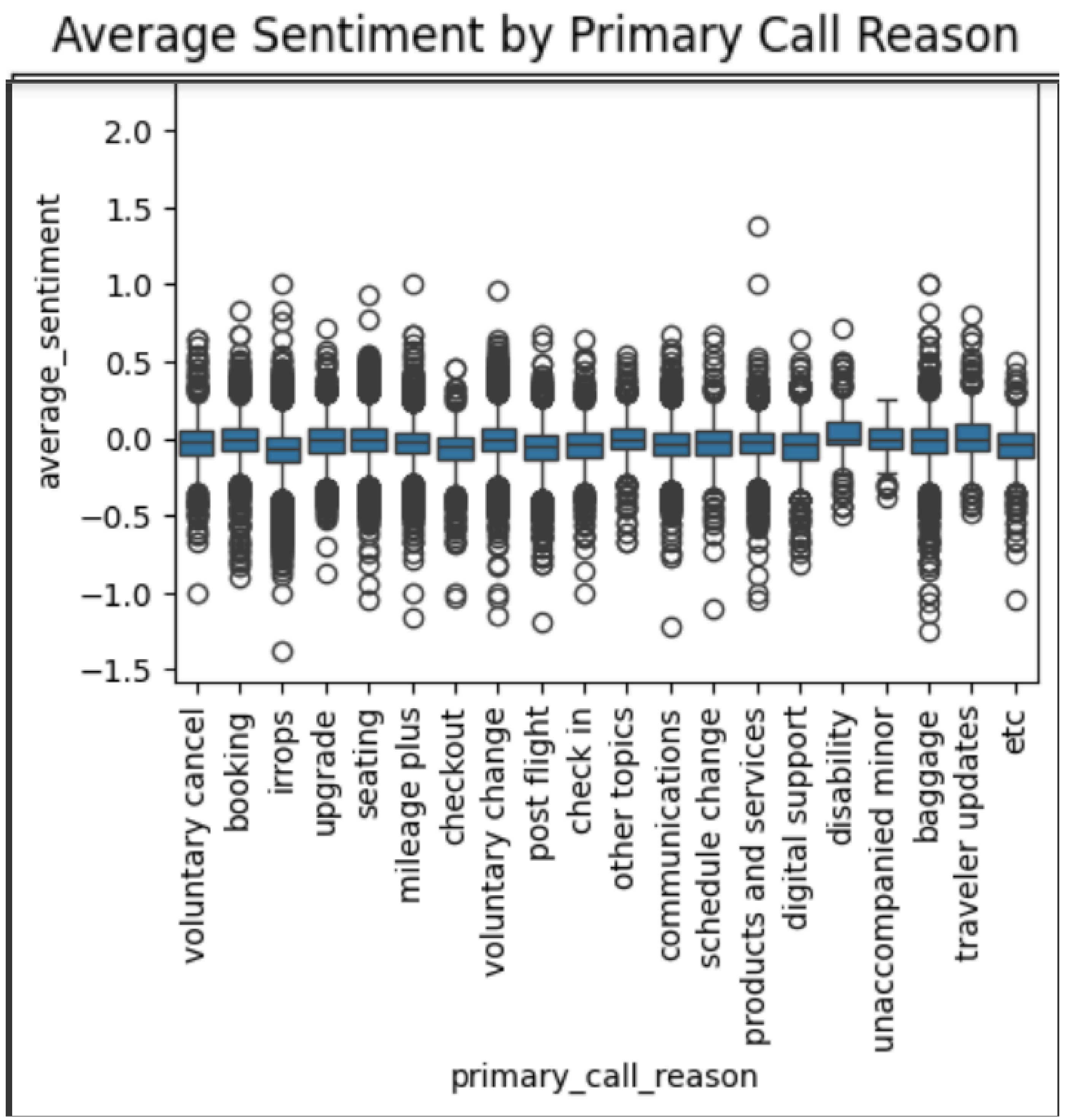
4. Evaluation

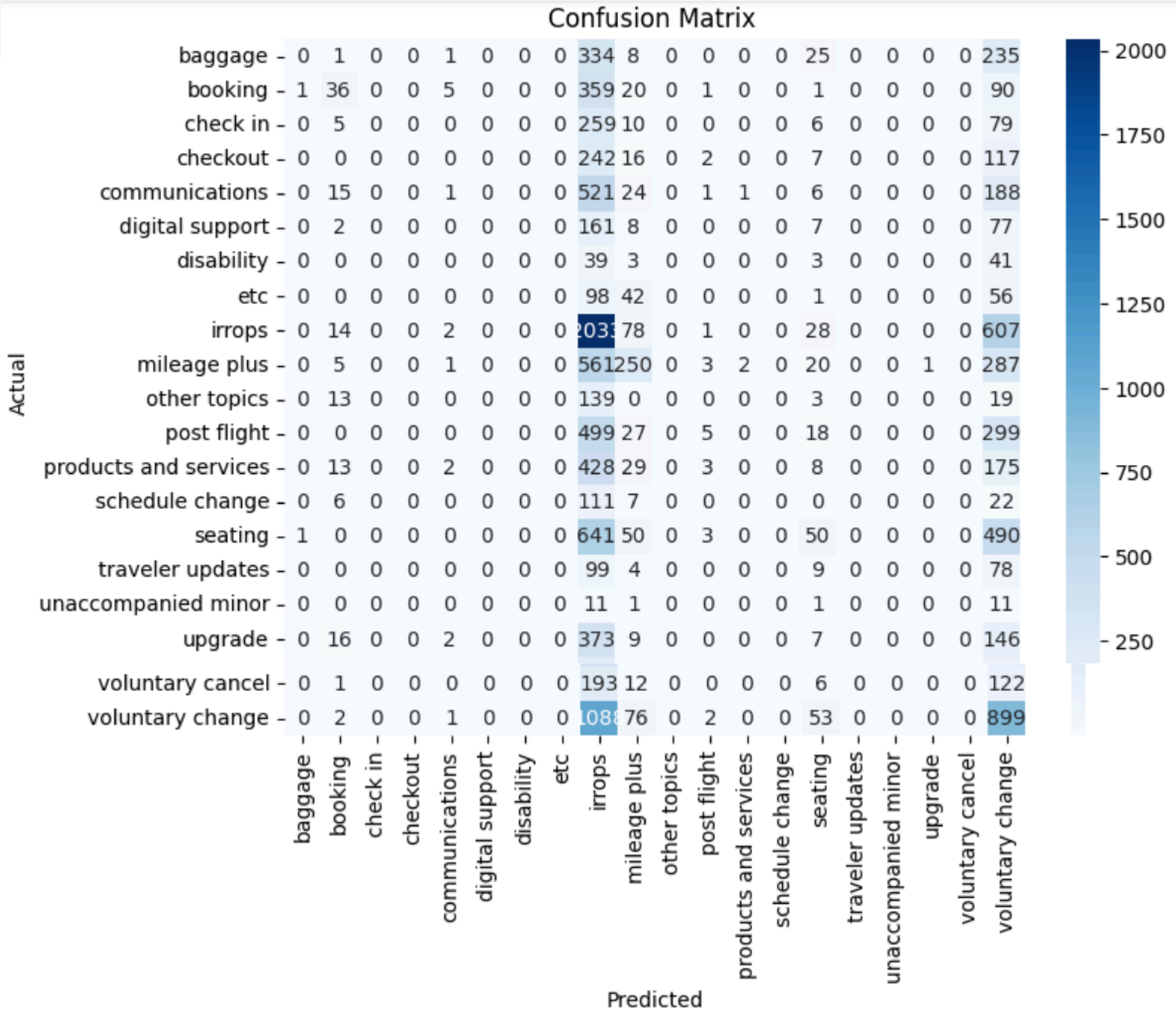
- Assessed performance using precision, recall, F1-score, and confusion matrix.

5. Test Prediction

- Made predictions on the test set and saved results to test_output.csv.


Distribution of Primary Call Reasons
count
12000
10000
8000
6000
4000
2000
0
irrops
voluntary change
seating
mileage plus
post flight
communications
products and services
baggage
upgrade
booking
check in
checkout
voluntary cancel
digital support
etc
traveler updates
other topics
schedule change
disability
unaccompanied minor


Average Sentiment by Primary Call Reason
average_sentiment
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
voluntary cancel
booking
irrops
upgrade
seating
mileage plus
checkout
voluntary change
post flight
check in
other topics
communications
schedule change
products and services
digital support
disability
unaccompanied minor
baggage
traveler updates
etc
primary_call_reason



Confusion Matrix
Actual
Predicted
baggage 0 1 0 0 1 0 0 0 334 8 0 0 0 0 25 0 0 0 0 235
booking 1 36 0 0 5 0 0 0 359 20 0 1 0 0 1 0 0 0 0 90
check in 0 5 0 0 0 0 0 0 259 10 0 0 0 0 6 0 0 0 0 79
checkout 0 0 0 0 0 0 0 0 242 16 0 2 0 0 7 0 0 0 0 117
communications 0 15 0 0 1 0 0 0 521 24 0 1 1 0 6 0 0 0 0 188
digital support 0 2 0 0 0 0 0 0 161 8 0 0 0 0 7 0 0 0 0 77
disability 0 0 0 0 0 0 0 0 39 3 0 0 0 0 3 0 0 0 0 41
etc 0 0 0 0 0 0 0 0 98 42 0 0 0 0 1 0 0 0 0 56
irrops 0 14 0 0 2 0 0 0 2033 78 0 1 0 0 28 0 0 0 0 607
mileage plus 0 5 0 0 1 0 0 0 561 250 0 3 2 0 20 0 0 1 0 287
other topics 0 13 0 0 0 0 0 0 139 0 0 0 0 0 3 0 0 0 0 19
post flight 0 0 0 0 0 0 0 0 499 27 0 5 0 0 18 0 0 0 0 299
products and services 0 13 0 0 2 0 0 0 428 29 0 3 0 0 8 0 0 0 0 175
schedule change 0 6 0 0 0 0 0 0 111 7 0 0 0 0 0 0 0 0 0 22
seating 1 0 0 0 0 0 0 0 641 50 0 3 0 0 50 0 0 0 0 490
traveler updates 0 0 0 0 0 0 0 0 99 4 0 0 0 0 9 0 0 0 0 78
unaccompanied minor 0 0 0 0 0 0 0 0 11 1 0 0 0 0 1 0 0 0 0 11
upgrade 0 16 0 0 2 0 0 0 373 9 0 0 0 0 7 0 0 0 0 146
voluntary cancel 0 1 0 0 0 0 0 0 193 12 0 0 0 0 6 0 0 0 0 122
voluntary change 0 2 0 0 1 0 0 0 1088 76 0 2 0 0 53 0 0 0 0 899
baggage
booking
check in
checkout
communications
digital support
disability
etc
irrops
mileage plus
other topics
post flight
products and services
schedule change
seating
traveler updates
unaccompanied minor
upgrade
voluntary cancel
voluntary change
2000
1750
1500
1250
1000
750
500
250

Classification metrics

	precision	recall	f1-score	support
baggage	0.00	0.00	0.00	604
booking	0.28	0.07	0.11	513
check in	0.00	0.00	0.00	359
checkout	0.00	0.00	0.00	384
communications	0.07	0.00	0.00	757
digital support	0.00	0.00	0.00	255
disability	0.00	0.00	0.00	86
etc	0.00	0.00	0.00	197
irrops	0.25	0.74	0.37	2763
mileage plus	0.37	0.22	0.28	1130
other topics	0.00	0.00	0.00	174
post flight	0.24	0.01	0.01	848
products and services	0.00	0.00	0.00	658
schedule change	0.00	0.00	0.00	146
seating	0.19	0.04	0.07	1235
traveler updates	0.00	0.00	0.00	190
unaccompanied minor	0.00	0.00	0.00	24
upgrade	0.00	0.00	0.00	553
voluntary cancel	0.00	0.00	0.00	334
voluntary change	0.22	0.42	0.29	2121
accuracy			0.25	13331
macro avg	0.08	0.07	0.06	13331
weighted avg	0.17	0.25	0.16	13331

Model Accuracy: 0.24559297877128497

UNITED

FINAL OUTPUT

Made predictions on the test set and saved results to test_shohel.csv.

	call_id	primary_call_reason
0	7732610078	seating
1	2400299738	baggage
2	6533095063	booking
3	7774450920	baggage
4	9214147168	baggage
...	...	...
5152	5300201106	baggage
5153	727694488	baggage
5154	147487837	irrops
5155	5330794838	seating
5156	8332067080	baggage

5157 rows × 2 columns

UNITED

THANK YOU

FOR YOUR ATTENTION