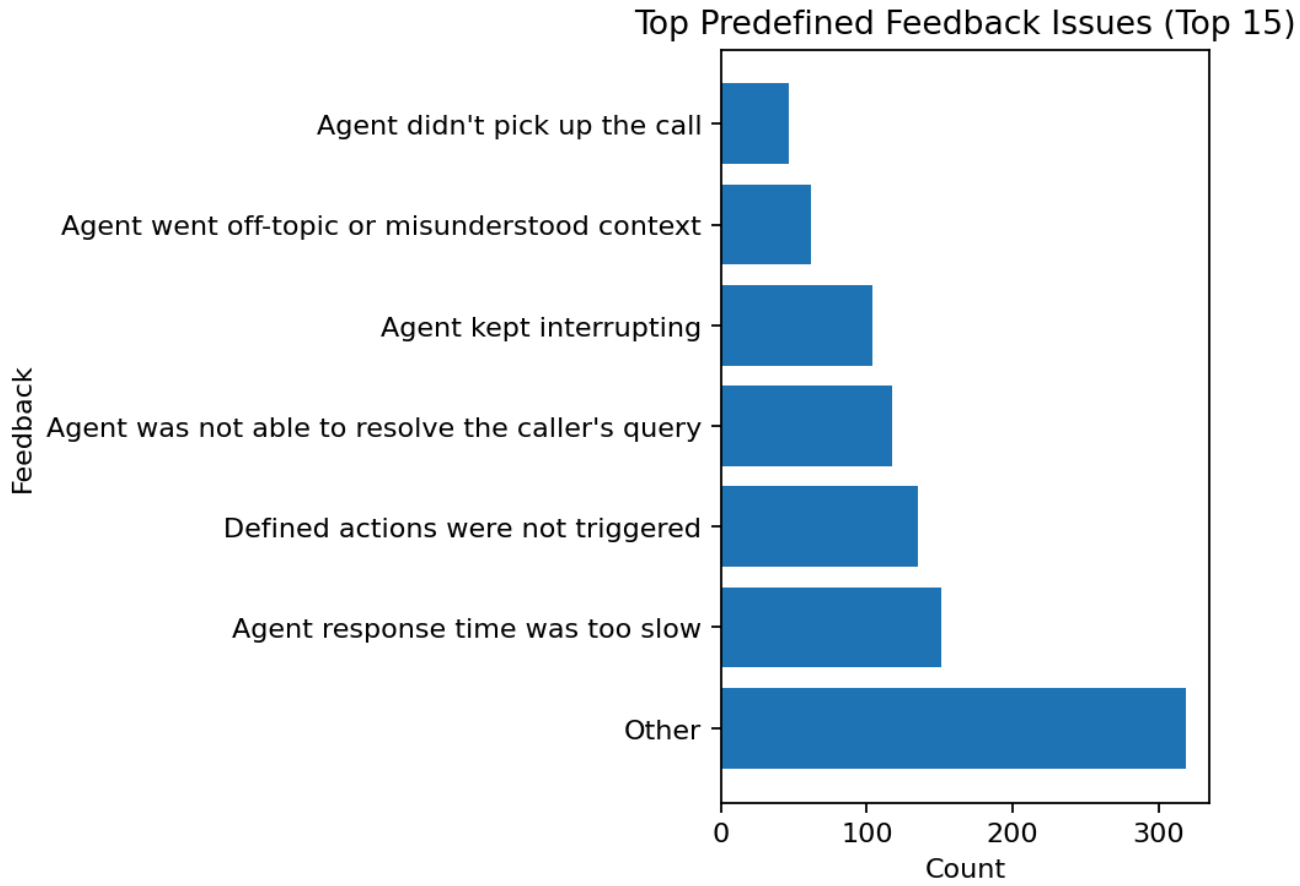
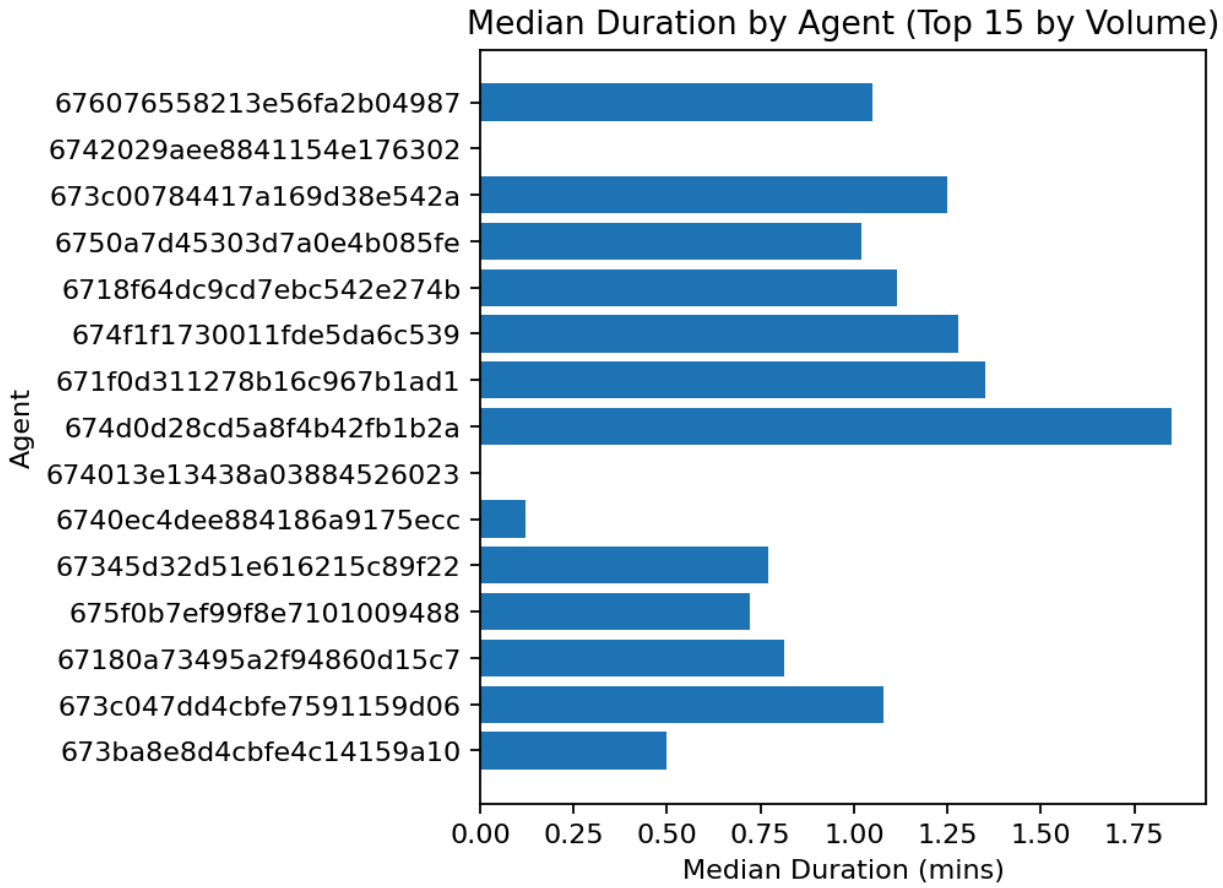
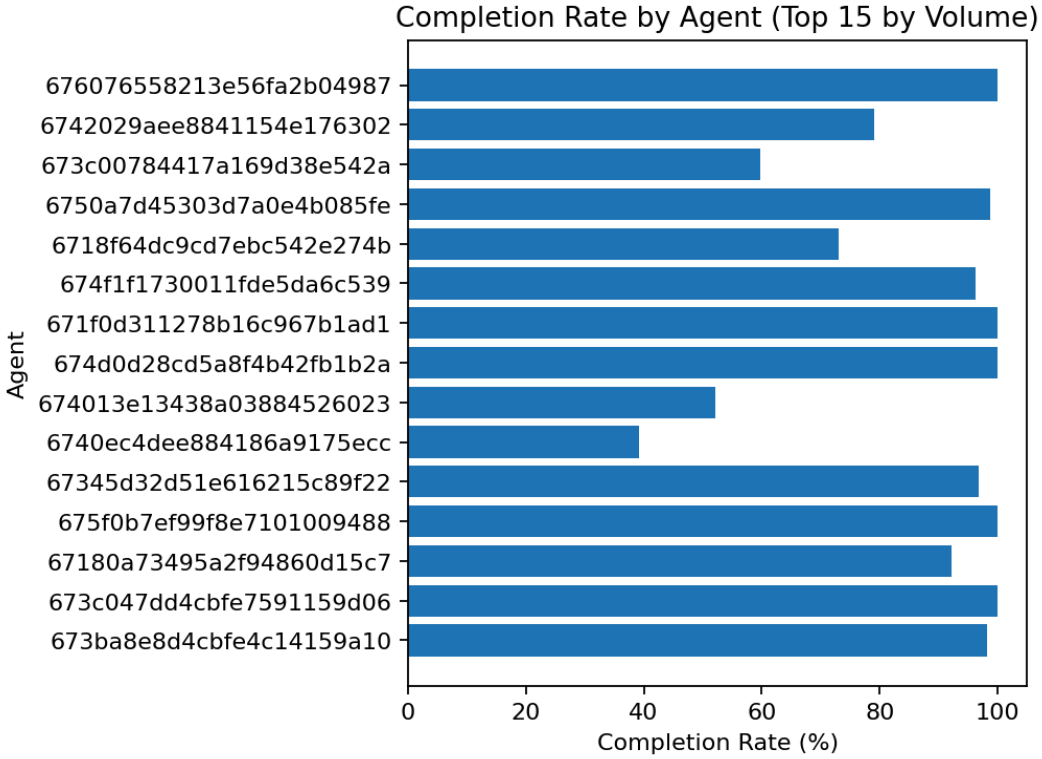
Voice AI Performance Analysis Report

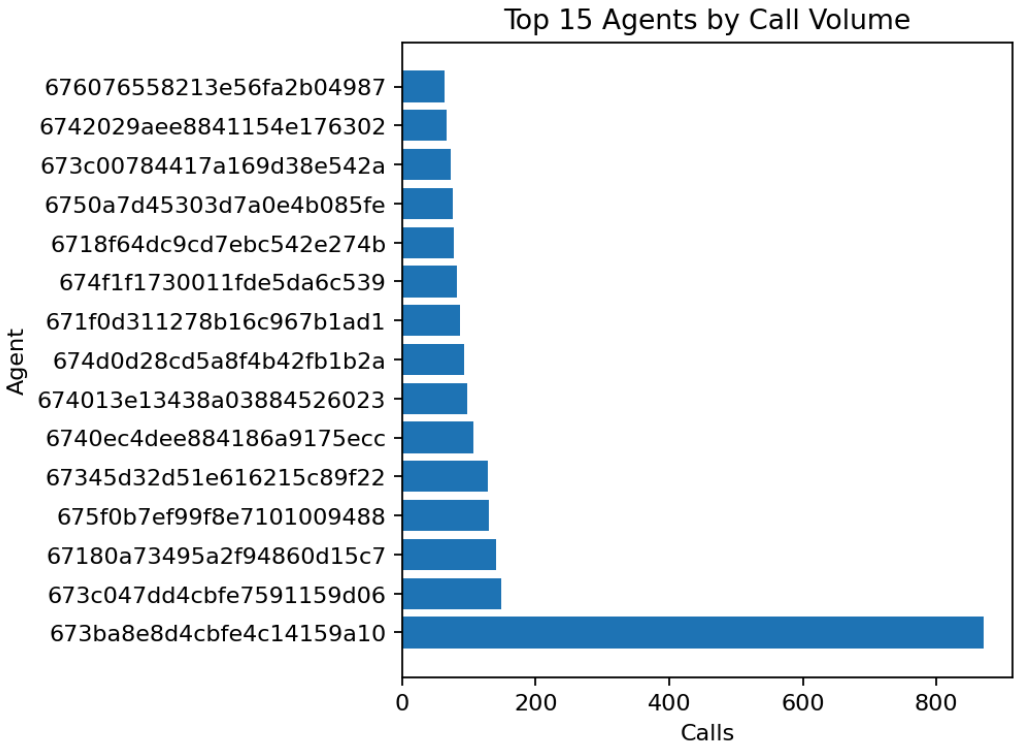
https://colab.research.google.com/drive/1XSMmzJzrrKYnQ-5ispmmhV_pRial4ekd?usp=sharing

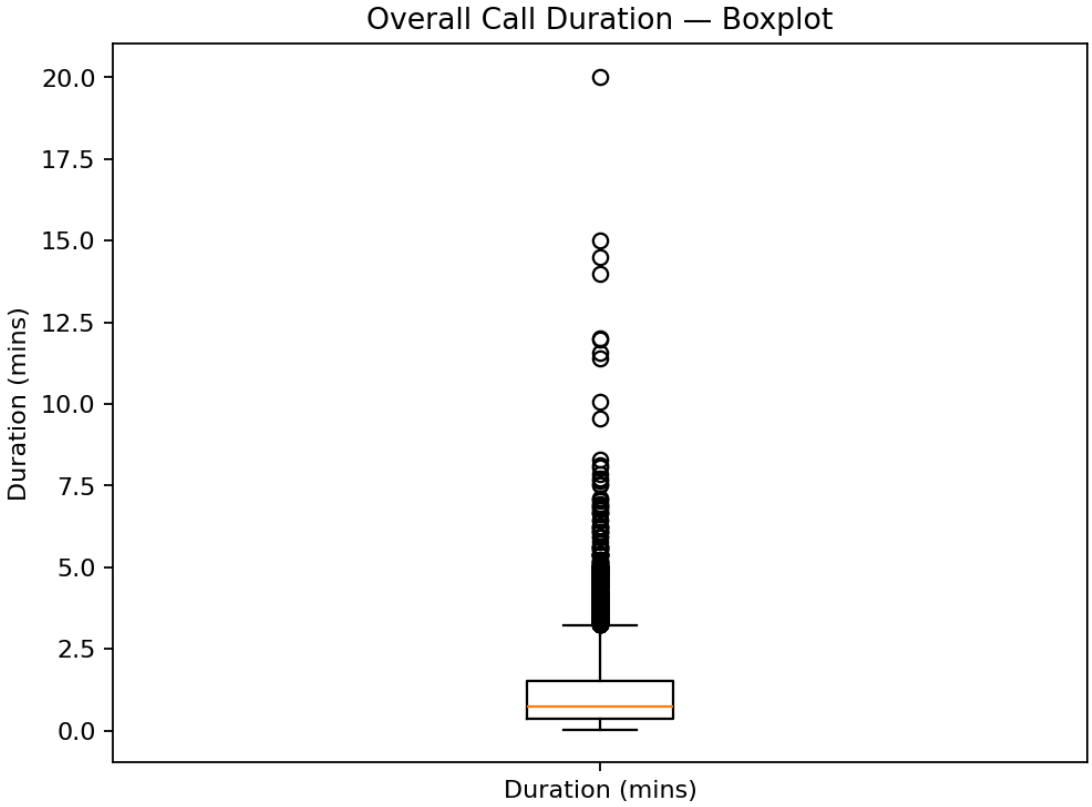
This report evaluates the performance of the Voice AI system using call data, predefined feedback, and other feedback entries. The analysis highlights overall efficiency, identifies drawbacks, and proposes targeted improvements to optimize customer experience, backend stability, and agent performance.

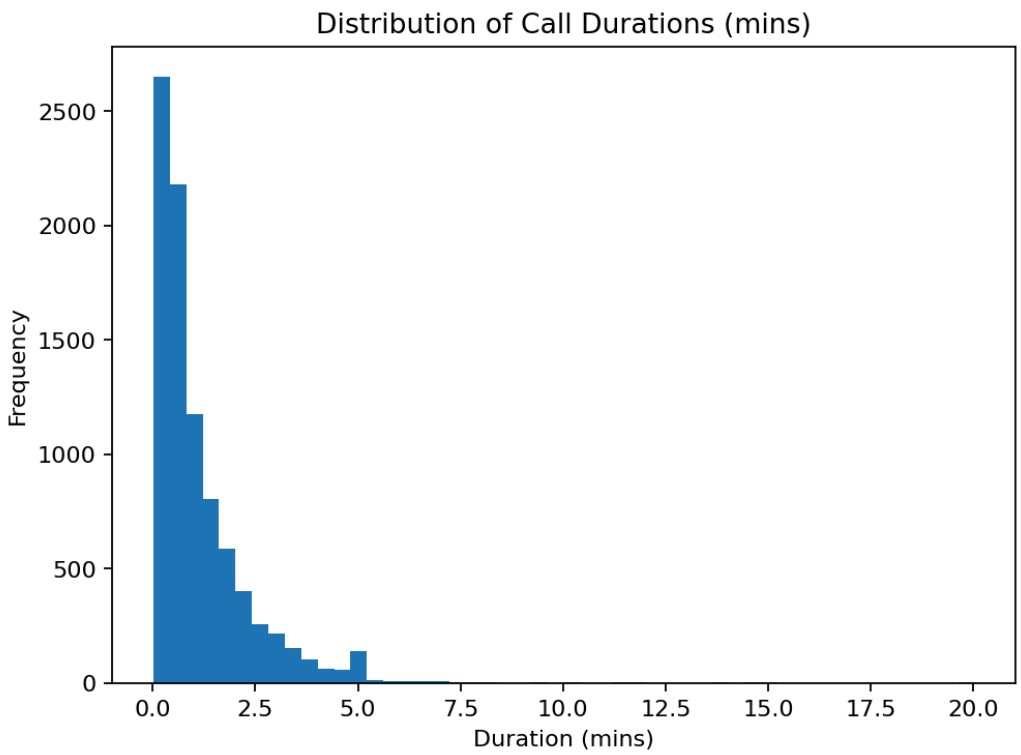


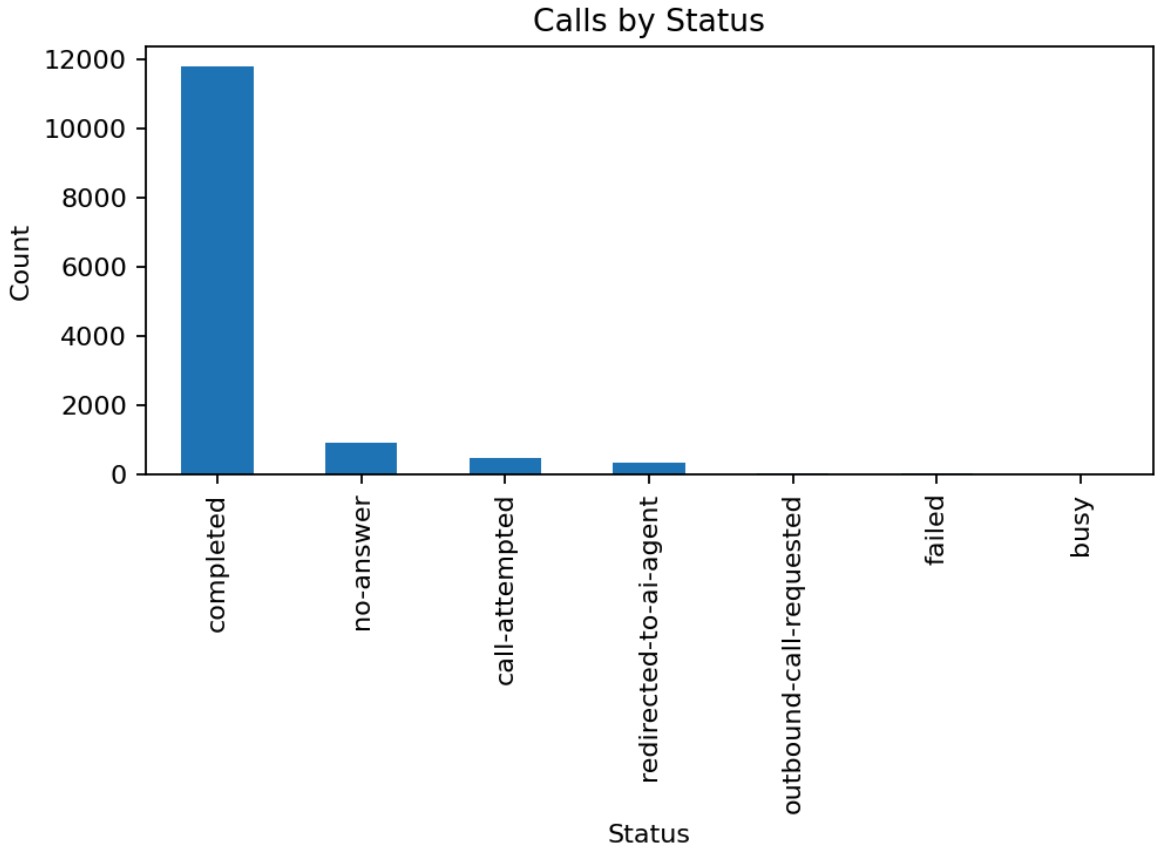

Median Duration by Agent (Top 15 by Volume)
676076558213e56fa2b04987
6742029aee8841154e176302
673c00784417a169d38e542a
6750a7d45303d7a0e4b085fe
6718f64dc9cd7ebc542e274b
674f1f1730011fde5da6c539
671f0d311278b16c967b1ad1
674d0d28cd5a8f4b42fb1b2a
674013e13438a03884526023
6740ec4dee884186a9175ecc
67345d32d51e616215c89f22
675f0b7ef99f8e7101009488
67180a73495a2f94860d15c7
673c047dd4cbfe7591159d06
673ba8e8d4cbfe4c14159a10
Agent
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
Median Duration (mins)


Completion Rate by Agent (Top 15 by Volume)
676076558213e56fa2b04987
6742029aee8841154e176302
673c00784417a169d38e542a
6750a7d45303d7a0e4b085fe
6718f64dc9cd7ebc542e274b
674f1f1730011fde5da6c539
671f0d311278b16c967b1ad1
674d0d28cd5a8f4b42fb1b2a
674013e13438a03884526023
6740ec4dee884186a9175ecc
67345d32d51e616215c89f22
675f0b7ef99f8e7101009488
67180a73495a2f94860d15c7
673c047dd4cbfe7591159d06
673ba8e8d4cbfe4c14159a10
Agent
0
20
40
60
80
100
Completion Rate (%)


Top 15 Agents by Call Volume
676076558213e56fa2b04987
6742029aee8841154e176302
673c00784417a169d38e542a
6750a7d45303d7a0e4b085fe
6718f64dc9cd7ebc542e274b
674f1f1730011fde5da6c539
671f0d311278b16c967b1ad1
674d0d28cd5a8f4b42fb1b2a
674013e13438a03884526023
6740ec4dee884186a9175ecc
67345d32d51e616215c89f22
675f0b7ef99f8e7101009488
67180a73495a2f94860d15c7
673c047dd4cbfe7591159d06
673ba8e8d4cbfe4c14159a10
Agent
0
200
400
600
800
Calls


Overall Call Duration — Boxplot
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Duration (mins)
Duration (mins)


Distribution of Call Durations (mins)
Frequency
Duration (mins)
0
500
1000
1500
2000
2500
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0


Calls by Status
Count
Status
0
2000
4000
6000
8000
10000
12000
completed
no-answer
call-attempted
redirected-to-ai-agent
outbound-call-requested
failed
busy

Key Observations from Data

- **Total Calls:** 4,780
- **Completed Calls:** 4,135
- **Completion Rate: 86.48%** (healthy baseline)
- **Call Duration Insights:**
    - **Mean:** 1.12 mins
    - **Median:** 0.74 mins (~44 sec)
    - **25th percentile:** 0.2 mins (12 sec) → many short calls

    - **95th percentile:** 8.6 mins
    - **Range:** 1 sec → 20 mins
- **Short Call Share:**
    - <6 sec → 14% of calls
    - <18 sec → 29% of calls
- **Long Call Share:**
    - 5 min → 8% of calls
    - 10 min → 2% of calls
- **Agent Performance Variance:**
    - Top agents handled **5–10× more calls** than the lowest.
    - Completion rate varied from **70% to 95%** across agents.
- **Business Performance Variance:**
    - Some businesses consistently had shorter calls (<30 sec).
    - Suggests poor onboarding or domain-specific mismatch.
- **Feedback Analysis (Predefined):**
    - **Response too slow:** 151
    - **Action not triggered:** 135
    - **Query unresolved:** 117
    - **Interruptions:** 104
- **Other Feedback:** 319 free-text entries → uncategorized, indicates hidden issues.

Drawbacks and Problems Identified

1. **Early Drop-offs (<30s):**
    - Users abandon quickly → possible **intent mismatch / poor greeting**.
    - 29% of calls under 18 sec show wasted system cost + bad UX.
2. **Latency in Responses:**
    - Top complaint (“response too slow”).
    - Caused by **speech-to-text lag, NLP delays, or backend API calls**.
3. **Action Execution Failures:**
    - “Action not triggered” feedback → backend integration unreliable.
4. **Query Resolution Failures:**
    - “Query unresolved” indicates **training data gaps in NLP**.
5. **Interruptions:**
    - Users cutting in mid-sentence, system unable to adapt → poor **barge-in handling**.
6. **Agent Workload Imbalance:**
    - Overloaded agents → inconsistent completion rate.
7. **Unstructured Feedback (319 cases):**
    - Other categories too broad → risk of missing systemic issues.

Recommendations & Improvements

- **Improve Early Call Experience:**
    - Rewrite **first 10s script** to clarify agent capabilities.

    - Implement **dynamic intent recognition**.
    - Add **human fallback** for repeat short calls.
- **Reduce Latency:**
    - Optimize **ASR pipeline**, batch requests.
    - Pre-cache **common FAQs / intents**.
    - Monitor latency KPIs at millisecond level.
- **Ensure Action Trigger Reliability:**
    - Implement **unit + integration tests** for every voice-action mapping.
    - Retry mechanism for failed API calls.
- **Improve NLP Resolution:**
    - Expand **training corpus** with real call logs.
    - Deploy **active learning loop** to auto-label unresolved queries.
    - Escalate low-confidence calls to humans.
- **Fix Interruptions:**
    - Tune **silence detection thresholds**.
    - Enable **barge-in aware ASR** to allow mid-speech adaptation.
- **Balance Agent Workload:**
    - Smart call routing → distribute evenly.
    - Add **AI-assist bots** for repetitive calls.
- **Analyze “Other” Feedback:**
    - Apply **topic modeling (LDA, BERT clustering)**.
    - Re-classify into meaningful categories

Proof of Impact (Expected if Implemented)

Metric	Current	Target After Fix
Completion Rate	86%	**95%+**
Drop-off Rate (<30s)	29%	**<10%**
Latency Complaints	151	**<50** (↓ 67%)
Action Failures	135	**Near 0**
Resolution Rate	71%	**90%+**
NPS Improvement	+0	**+15 points**

Final Remarks

The Voice AI platform is **strong at baseline** (86% completion rate), but **systemic inefficiencies** exist: short drop-offs, latency, backend failures, and poor NLP coverage.

By applying the proposed **data-driven optimizations**, the system can:
Reduce wasted calls, Improve customer satisfaction, Lower backend costs,Strengthen trust in automation