

Urban Mobility and Transport Insights in Greater Boston

Iteration Submission: Project Plan and Readiness

Eric Fu — Dhanrithii Deepa — Yuchen Cai

October 22, 2025

Abstract

We propose a composite **Rider Satisfaction Score (RSS)** for Greater Boston transit that blends ridership demand, service performance, and demographic equity indicators. This document answers all prompt questions: goals and scope, phased deliverables, milestones, capability assessment, data plan, team roles, skills and tools, initial setup, and submission artifacts. Team roles: **Eric Fu** (Database Management), **Dhanrithii Deepa** (Data Visualization), **Yuchen Cai** (Advanced Analytics & Insights).

1. Project Kickoff

1.1 Specific Goals and Expected Outcomes

Goal. Design a reproducible, interpretable **RSS** computed at multiple granularities (system, line, station, route–time window) to surface mobility pain points and equity gaps.

1.2 Overall Project Outcomes

1) *Satisfaction Metric Framework*

A clear, reproducible framework for combining multiple factors (e.g., usage, service quality, community context) into an overall satisfaction indicator, applicable at different levels of detail.

2) *Data Assets*

- Curated input datasets with brief source notes.
- A simple data dictionary describing each metric and how it is prepared.
- A configuration file (or section) documenting parameters and assumptions.

3) *Reproducible Workflow*

- A basic, repeatable process that ingests data, performs checks, and produces the satisfaction outputs.
- Minimal quality checks to flag missing or unusual values.

4) *Visuals & Insights*

- A small set of clear charts that illustrate results and trends.
- A short written summary highlighting key takeaways and potential implications.

5) *Validation & Sensitivity*

- Simple sensitivity checks to show how results change under reasonable parameter adjustments.
- Basic plausibility checks against known events or expectations.

6) *Delivery & Documentation*

- A concise report summarizing scope, approach, roles, tools, and findings.
- A repository containing the report, workflow files, brief instructions, and a progress tracker.

7) *Success Criteria*

- Coverage: results are produced consistently for the selected pilot scope.
- Clarity: methods and assumptions are easy to understand and review.
- Actionability: visuals and summaries help inform next steps or decisions.
- Reproducibility: running the documented steps consistently recreates the results.

1.3 Clear Project Scope (This Iteration)

In scope: Descriptive analytics, pilot RSS on limited time window/lines, equity gap readout, baseline charts or dashboard stub.

Out of scope: Causal inference, forecasting, simulation/optimization, productionized pipelines.

1.4 Key Deliverables by Phase

- **Phase A — Framing:** Problem statement, KPIs, risks, indicator list.
- **Phase B — Data/Environment:** Repo/env setup; minimal ETL; data dictionary.
- **Phase C — Analytics/Viz:** RSS v0.1; sensitivity; equity readout; charts.
- **Phase D — Package:** Overleaf PDF (3 pages); Excel tracker; GitHub updated.

1.5 Milestones and Deadlines

Date	Milestone	Definition of Done
Oct 25, 2025	Scope lock	Questions, KPIs, risks, indicators finalized.
Oct 28, 2025	Env ready	Repo, <code>README</code> , <code>requirements.txt</code> / <code>env.yml</code> .
Nov 1, 2025	Pilot data loaded	Ingested pilot slice; data dictionary draft.
Nov 6, 2025	RSS v0.1 & viz	Score computed; 3–5 charts; sensitivity table.
Nov 10, 2025	Submission package	PDF + Excel tracker + GitHub tagged for review.

2. Team Discussions

2.1 Core Skills by Member

- **Eric Fu** — schema design, ingestion/ETL, SQL views, data QA.
- **Dhanrithii Deepa** — visual encodings, dashboards, narrative captions.
- **Yuchen Cai** — indicator math, normalization/weights, sensitivity/equity analysis.

2.2 How Expertise Maps to Tasks

Owner	This-iteration responsibilities
Eric	Ingest pilot data; SQLite (or Postgres) setup; QA checks; SQL views feeding RSS inputs.
Dhanrithii	Altair/Matplotlib chart pipeline; station/line heatmaps; factor contribution bars; dashboard stub.
Yuchen	Define indicators/weights; compute RSS; equity gap readout; sensitivity analysis.

2.3 Missing Skills and Risks

Geospatial joins, GTFS/performance nuances. **Plan:** 1–2 hr spike with *GeoPandas* on a toy join; pair programming on first map; keep transformations under version control.

2.4 Tools: Experienced vs To Learn

Experienced: Python, pandas/NumPy, SQL, Matplotlib/Altair, Git/GitHub, Overleaf.

To learn (lightweight): GeoPandas/Folium; optional Great Expectations; optional Streamlit.

2.5 Languages and Platforms

Languages: Python + SQL. **Storage (prototype):** CSV → **SQLite** (or Postgres if needed).

Viz: Altair primary; Matplotlib fallback. **Optional UI:** Streamlit stub.

3. Skills and Tools Assessment

3.1 External Resources

Faculty/TA office hours; official docs (MBTA developer/GTFS, Bluebikes portal, ACS); peer consultation.

3.2 Fit-for-Purpose Stack

- **ETL/EDA:** Python (pandas, numpy), **pyarrow** for parquet.

- **Geospatial:** GeoPandas (& shapely); Folium (optional maps).
- **Validation (optional):** Great Expectations or unit checks in notebooks.
- **Viz:** Altair (declarative) + Matplotlib.
- **Repro:** requirements.txt/environment.yml; simple run script or Makefile.
- **Config:** weights.yml for indicators/weights/directions.

3.3 Role Assignment & Clarity

Tasks are mapped to strengths (Section 2). Each task has an *owner*, due date, and *Definition of Done*. Progress tracked in Excel and via GitHub issues/PRs.

4. Initial Setup

4.1 Development Environment

Python 3.10+ with: pandas, numpy, pyarrow, geopandas, matplotlib, altair, sqlalchemy, sqlite3 (or psycopg for Postgres); optional great_expectations, streamlit.

4.2 Version Control

GitHub repository created; all members have write access. **Branching:** main (protected) + short-lived feature branches. Require PR + 1 review before merge to main.

4.3 Repository Links & Files

GitHub Repository: <https://github.com/ORG/REPO> (replace with actual)

4.4 Public Datasets Used (Links & Titles)

- MBTA Rapid Transit Speed Restrictions by Day
<https://mbta-massdot.opendata.arcgis.com/datasets/d73ed67e4cc84a84b818ea2c5caef696/about>
- MBTA The RIDE Ridership
https://mbta-massdot.opendata.arcgis.com/datasets/e93e4e4820ca4719b3c4134ae0865053_0/explore
- MBTA Rapid Transit Travel Times 2024
<https://mbta-massdot.opendata.arcgis.com/datasets/0b4dc16b8b984836962229865d5b573b/about>
- MBTA Blue Book Open Data Portal (tag: 2024)
<https://mbta-massdot.opendata.arcgis.com/search?tags=2024>
- MBTA Monthly Ridership By Mode and Line
https://mbta-massdot.opendata.arcgis.com/datasets/2048258a18354256a650d41f8fe4532c_0/explore
- MBTA 2024 System-Wide Passenger Survey
https://mbta-massdot.opendata.arcgis.com/datasets/7da1f62034f64cb4bc9e2afefe9a1fdc_0/explore
- Fall 2024 MBTA Rail Ridership by SDP Time Period, Route/Line, and Stop
https://mbta-massdot.opendata.arcgis.com/datasets/d4610a65064a4d3c8536c75d520e0012_0/explore