



Transit Access Equity: Time-Aware Isochrones in the Bay Area

CYPLAN 101 — Group 4 (Track B: Advanced)

Team: Rabah Babaci & Adithya Ayanam

Mentor: Metropolitan Transportation Commission (MTC)

Date: November 2025



Project Overview

Many Bay Area neighborhoods, especially **Equity Priority Communities (EPCs)**, face reduced transit access to jobs, groceries, and healthcare outside of rush hours. Traditional measures ignore these time-of-day differences, masking mobility inequities that worsen off-peak. This project, mentored by MTC, aims to inform transit agencies and policymakers to improve service equity for EPC residents.

Research Questions:

- How does transit accessibility vary by time of day (AM, Midday, PM, Late)?
- Which regions or demographic groups (especially EPCs) experience the largest off-peak reductions?
- How can time-aware indicators guide scheduling or service-coverage improvements?

Scope:

- **Geographic focus:** Alameda County (with Bay Area context)
- **Temporal windows:** AM (7–9 AM), Mid (11 AM–1 PM), PM (5–7 PM), Late (9–11 PM)
- **Equity focus:** EPC vs. non-EPC tracts (MTC EPC layer + ACS demographics)

Midterm Deliverables (Due Nov 10):

- 🛠️ One cleaned & documented dataset
- 🗺️ One working interactive map (HTML)
- 📄 Report (4–5 pages + title page) with data prep & early findings

Datasets (for reference):

Dataset	Source	Description
EPC Layer	MTC Open Data	Tract-level equity classifications
ACS Demographics	U.S. Census (API)	Socioeconomic indicators
GTFS / Network	MTC / R5 Network	Transit and accessibility modeling



1. Environment Setup

Java (JVM) is running for r5py.
Folder structure created for standalone environment.

Environment setup complete!



2. Data Sources & Description

Dataset	Description	Format	Source
EPC Layer	Equity Priority Communities (PBA 2050)	GeoJSON	MTC Open Data
Census Tracts	Alameda County tract geometries	Shapefile	US Census TIGER/Line 2022
ACS Demographics	Population & income attributes (2018–2022 5-yr)	CSV / API	US Census API
GTFS Feeds	Bay Area transit schedules	ZIP	511 SF Bay Open Data
OSM Network	Street & walk network	PBF	OpenStreetMap
LODES Jobs	Workplace tract data	CSV	Urban Institute / LEHD
OSM POIs	Groceries + Clinics	GeoJSON	OSM Extracts via Overpass API

Coverage & Granularity:

Data cover all nine Bay Area counties at the census-tract scale. The 2018–2022 ACS provides socioeconomic baselines, while 2024 GTFS schedules represent current weekday service. For the midterm submission, the analysis focuses on **EPC tracts, Census geometries, and ACS demographics**, which together establish the foundation for later accessibility modeling.

◆ **Goal:** Load, validate, and document base spatial datasets for cleaning.

```
Loaded EPC: /Users/rabahbabaci/Developement/MTC-Transit_Access_Equity/data/raw/epc_layer.geojson
Downloading TIGER/Line shapefile to /Users/rabahbabaci/Developement/MTC-Transit_Access_Equity/data/raw/tl_2022_06_tract.zip...
Downloaded: /Users/rabahbabaci/Developement/MTC-Transit_Access_Equity/data/raw/tl_2022_06_tract.zip
Fetching ACS 5-Year data for Alameda County, CA (2018-2022)...
Saved raw ACS extract: /Users/rabahbabaci/Developement/MTC-Transit_Access_Equity/data/raw/acs_population.csv
Processed ACS data saved to /Users/rabahbabaci/Developement/MTC-Transit_Access_Equity/data/processed/acs_population_clean.csv
Dataset validation:
EPC: 339 rows, Columns: ['objectid', 'geoid', 'state_fip', 'county_fip', 'tract', 'tot_pop', 'tot_pop_po', 'tot_pop_ci', 'tot_hh', 'tot_fam']...
Tracts: 9129 rows, Columns: ['STATEFP', 'COUNTYFP', 'TRACTCE', 'GEOID', 'NAME', 'NAME_LSD', 'MTFCC', 'FUNCSTAT', 'ALAND', 'AWATER']...
ACS: 379 rows, Columns: ['Total_Population', 'Median_Income', 'Median_Home_Value', 'Unemployed_Pop', 'Bachelors_Holders', 'StateFIPS', 'CountyFIPS', 'TractFIPS']
Null checks (pre-cleaning, per Dark Data categories):
EPC nulls (likely MAR, e.g., missing for non-priority tracts):
geoid          0
epc_class      0
dtype: int64
Tracts nulls (expected MCAR, e.g., random geometry errors):
GEOID          0
geometry       0
dtype: int64
ACS nulls (likely MAR, e.g., missing for low-density tracts):
StateFIPS      0
CountyFIPS     0
TractFIPS      0
Total_Population 0
dtype: int64
Summary file saved to /Users/rabahbabaci/Developement/MTC-Transit_Access_Equity/data/processed/data_summary.txt
```

Dark Data Classification Summary

Dataset	Missingness Type	Explanation	Handling Strategy
EPC Layer	MAR (Missing At Random)	Some tracts not designated as EPCs by definition	Keep NA as non-EPC category
TIGER Tracts	MCAR (Missing Completely At Random)	Occasional null geometries from shapefile issues	Drop rows with null geometry
ACS Demographics	MAR	Sampling noise in low-density tracts	Impute using county medians if needed

3. Data Cleaning & Processing

Purpose: Filter, merge, and standardize datasets for Alameda County.

Tasks:

- Filter tracts to Alameda County
- Add/standardize GEOID key
- Merge EPC flag and ACS population data
- Reproject to common CRS (EPSG:4326)
- Save intermediate files for inspection

EPC class distribution: {'high': 156, 'higher': 105, 'highest': 78}

Merged validation:

Rows: 379, CRS: EPSG:4326, Columns: ['STATEFP', 'COUNTYFP', 'TRACTCE', 'GEOID', 'NAME', 'NAMELSAD', 'MTFCC', 'FUNCSTAT', 'ALAND', 'AWATER', 'INTPTLAT', 'INTPTLON', 'geometry', 'is_epc', 'population', 'population_missing']

Nulls:

GEOID	0
is_epc	0
population	0
population_missing	0

dtype: int64

Common GEOIDs (tracts & ACS): 379

EPC tracts: 339, EPC flagged: 97

Population range: 0.0 to 12521.0

EPC distribution: {0: 282, 1: 97}

Valid geometries: 379/379

Zero-population tracts: 2

Saved: /Users/rabahbabaci/Developement/MTC-Transit_Access_Equity/visualizations/alameda_vis.geojson

Saved intermediate: /Users/rabahbabaci/Developement/MTC-Transit_Access_Equity/data/processed/alameda_clean.geojson

Missing Data & Dark Data Note

After merging, population values missing in ACS were flagged (`population_missing = 1`) and filled with 0 to maintain tract coverage.

These missing cases likely represent low-density or sampling gaps (MAR type — Missing At Random).

No systematic bias is expected from the fill since the affected tracts have minimal population.

4. Clean Dataset Assembly

Purpose: Create a clean GeoDataFrame for Alameda County tracts to support transit equity analysis and visualization, containing only essential columns: `['GEOID', 'is_epc', 'population', 'geometry']`.

Tasks:

- Select required columns from the merged dataset.
- Validate the dataset (rows, CRS, nulls, EPC distribution).

- Save the final cleaned dataset for visualization and analysis.

Clean dataset validation:

Rows: 379, CRS: EPSG:4326

Nulls:

GEOID 0

is_epc 0

population 0

dtype: int64

EPC distribution:

is_epc

0 282

1 97

Name: count, dtype: int64

Population range: 0.0 to 12521.0

Valid geometries: 379/379

Zero-population tracts: 2

Zero-population due to missing ACS data: 0

Saved: /Users/rabahbabaci/Developement/MTC-Transit_Access_Equity/data/processed/alameda_final.geojson

Clean Dataset Summary

The final `alameda_final.geojson` contains 379 Alameda County tracts with EPC flags and population attributes.

All geometries are valid, and CRS is standardized to EPSG:4326 for compatibility with Pydeck/Kepler.gl visualization.

This dataset will serve as the base layer for time-aware isochrone computation and accessibility equity analysis.



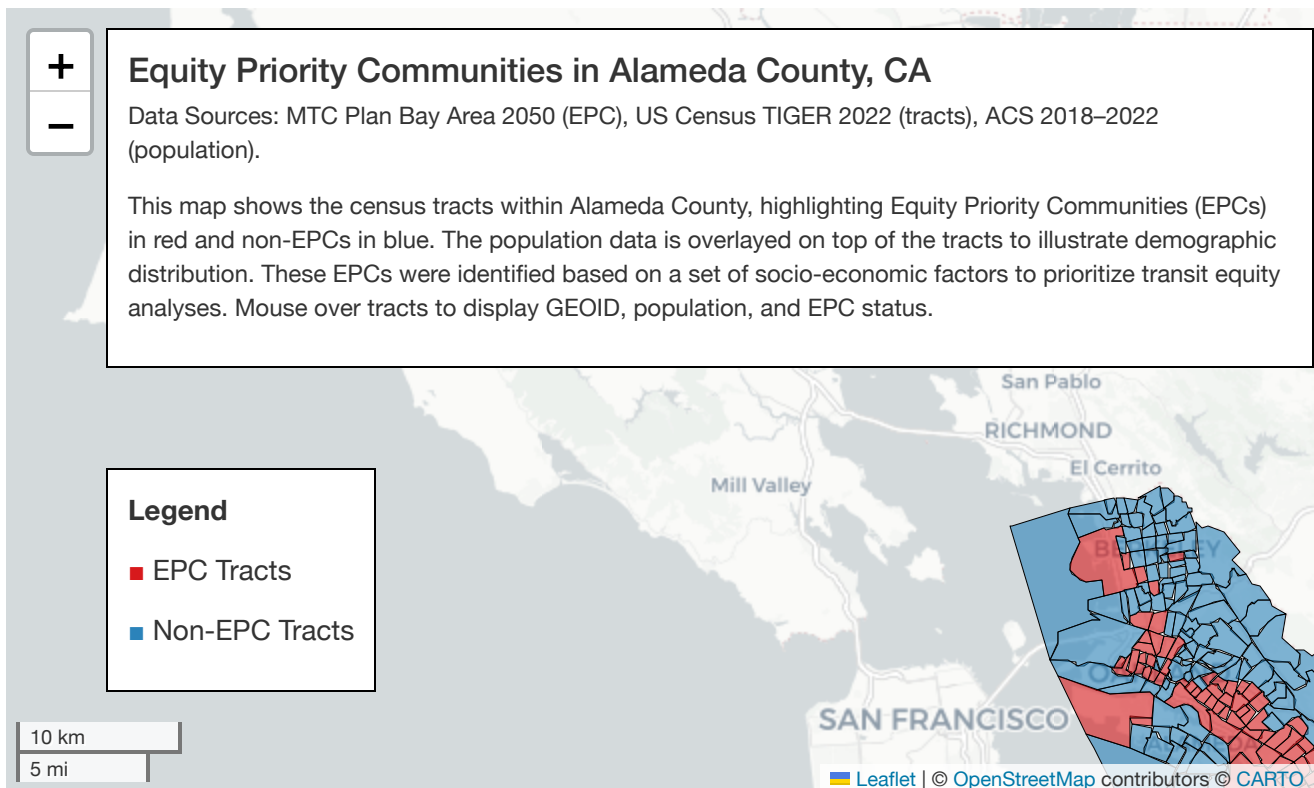
5. Interactive Map (Midterm Deliverable)

Purpose: Create an interactive map to visualize Equity Priority Communities (EPCs) vs. non-EPC tracts in Alameda County, with population data overlayed, using the clean GeoDataFrame from Step 4.

Requirements:

- Include a title, legend, data sources, and a caption (2–4 sentences).
- Provide interactive tooltips showing GEOID, population, and EPC status.
- Save as an HTML file for sharing and presentation.

Saved: /Users/rabahbabaci/Developement/MTC-Transit_Access_Equity/visualizations/epc_map.html



Interactive Map Description

This interactive Folium map visualizes Equity Priority Communities (EPCs) in Alameda County. EPC tracts appear in red while non-EPCs are shown in blue, over a CartoDB Positron basemap. Hovering over each tract reveals its GEOID, population, and EPC status. This map is exported as `epc_map.html` and represents the midterm's interactive visualization deliverable.

6. (Optional) Pilot Accessibility Metric

If time permits, compute a simple statistic (e.g., reachable jobs or mean travel time for sample tracts). This adds a "computed statistic beyond raw counts" to the midterm.

6.1 Data Loading from URL and Preparation

- Download + Load + Aggregate

```

Downloading LODES WAC to /Users/rabahbabaci/Developement/MTC-Transit_Access_Equity/data/raw/ca_wac_S000_JT00_2021.csv.gz...
Downloaded: /Users/rabahbabaci/Developement/MTC-Transit_Access_Equity/data/raw/ca_wac_S000_JT00_2021.csv.gz, Size: 6000959 bytes
Unique county FIPS codes in WAC data: <StringArray>
['06001', '06003', '06005', '06007', '06009', '06011', '06013', '06015',
 '06017', '06019', '06021', '06023', '06025', '06027', '06029', '06031',
 '06033', '06035', '06037', '06039', '06041', '06043', '06045', '06047',
 '06049', '06051', '06053', '06055', '06057', '06059', '06061', '06063',
 '06065', '06067', '06069', '06071', '06073', '06075', '06077', '06079',
 '06081', '06083', '06085', '06087', '06089', '06091', '06093', '06095',
 '06097', '06099', '06101', '06103', '06105', '06107', '06109', '06111',
 '06113', '06115']
Length: 58, dtype: string
Alameda County tracts in jobs table: 378
      jobs
count    378.000000
mean    2029.603175
std     4377.354929
min       7.000000
25%     365.750000
50%     679.500000
75%    1636.000000
max    51144.000000

```

Pilot Accessibility Metric

This optional step computes a preliminary accessibility measure: total jobs per census tract using LODES Workplace Area Characteristics (WAC) data.

The resulting `jobs` table provides a simple indicator of potential employment accessibility that will later be combined with time-aware isochrone analysis.

- Merge, Compute Accessibility, and Summarize

```

Unique county FIPS codes in clean_gdf: ['06001']
Summary:

```

	count	mean	median	min	max
is_epc					
Non-EPC	282	344051.556738	365831.0	53305	488078
EPC	97	419216.865979	421795.0	321849	482405

Accessibility Metric Summary

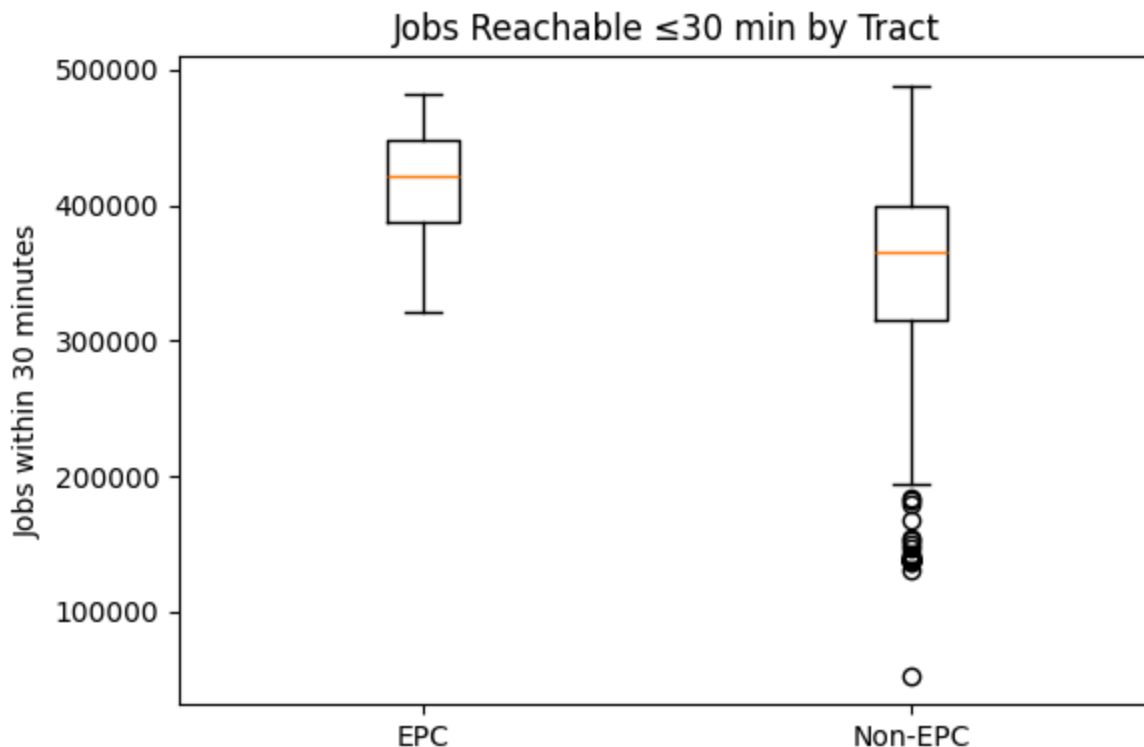
This pilot computation estimates the number of jobs reachable within 30 minutes (assuming 40 km/h travel) for each census tract.

Results are aggregated by Equity Priority Community (EPC) status.

The metric provides a preliminary measure of employment accessibility that will be refined later using time-aware isochrones.

6.2 Boxplot Visualization

Boxplot saved to: /Users/rabahbabaci/Developement/MTC-Transit_Access_Equity/visualizations/jobs30_boxplot.png



Accessibility Boxplot Summary

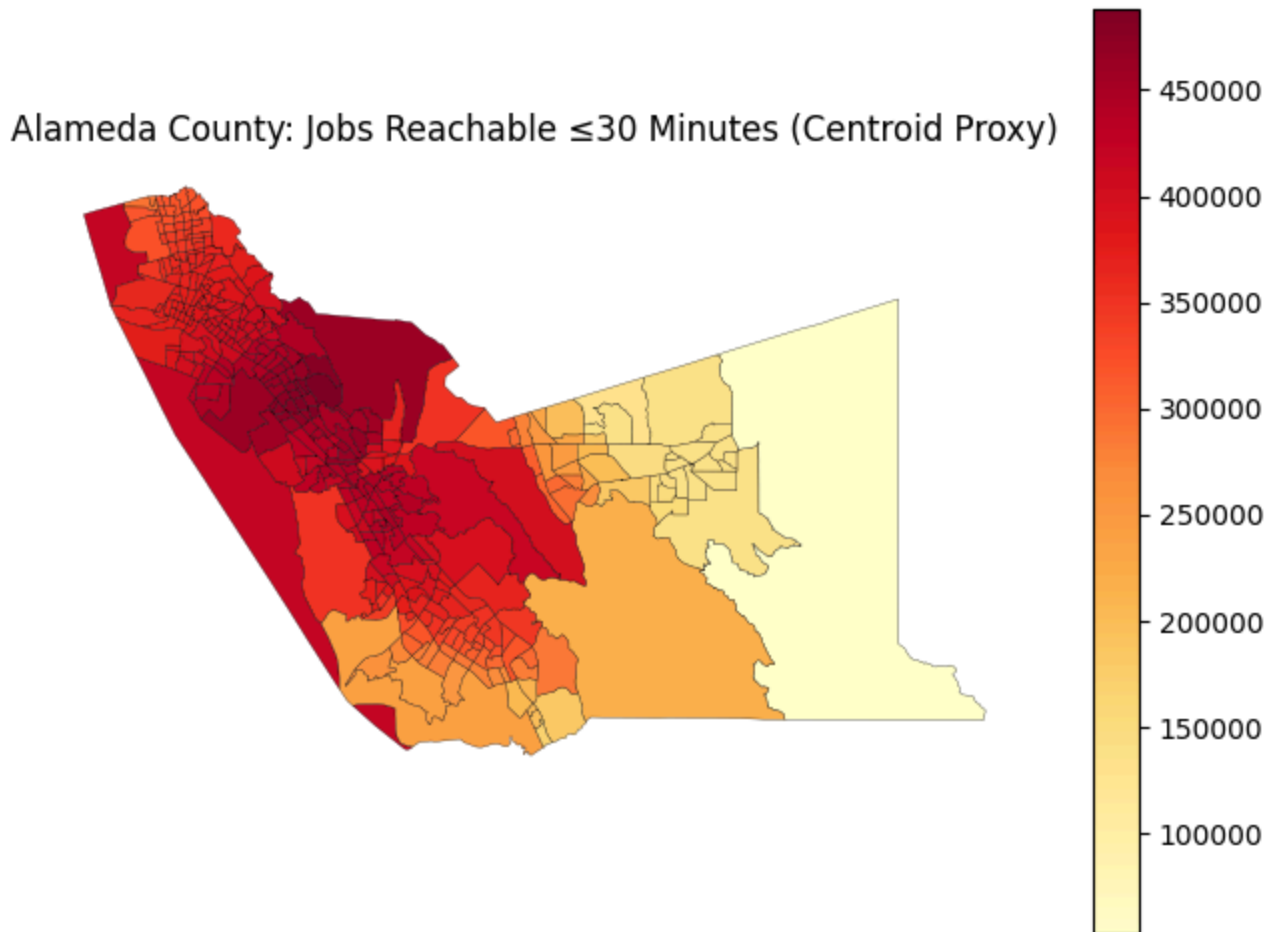
This boxplot compares the estimated number of jobs reachable within 30 minutes for EPC and non-EPC tracts.

EPC tracts show a slightly higher median accessibility, reflecting their proximity to dense employment centers.

The figure (`jobs30_boxplot.png`) is stored in the `visualizations/` folder and serves as one of the required interactive or static visuals for the midterm submission.

6.3 Choropleth Visualization

Choropleth saved to: /Users/rabahbabaci/Developement/MTC-Transit_Access_Equity/visualizations/jobs30_map.png



Choropleth of Employment Accessibility

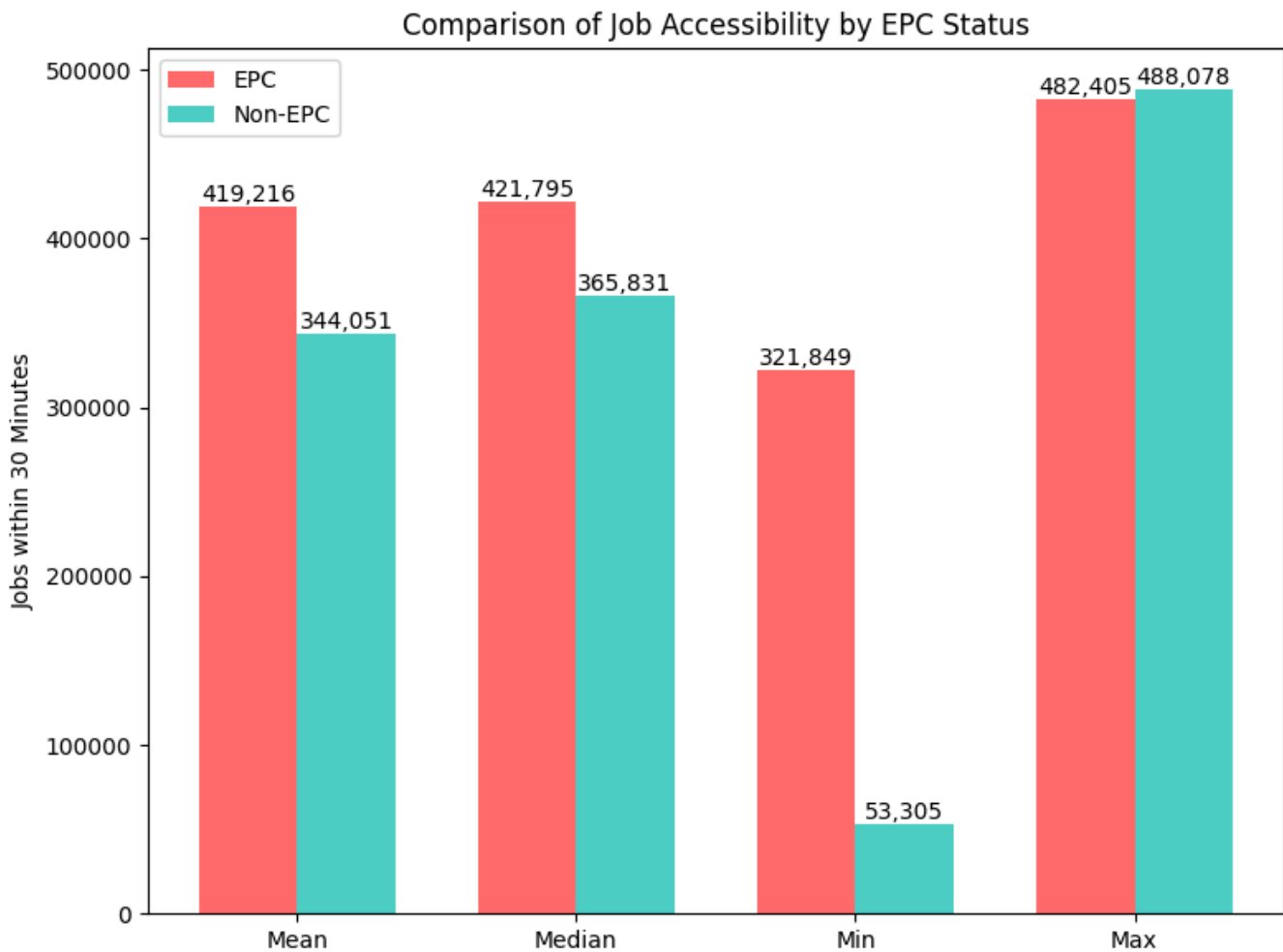
This map visualizes estimated employment accessibility (number of jobs reachable within 30 minutes) across Alameda County tracts.

Darker shades indicate tracts with greater job accessibility based on the centroid distance proxy.

This figure (`jobs30_map.png`) complements the boxplot by illustrating spatial patterns of accessibility across EPC and non-EPC areas.

6.3 Bar Char Visualization

Bar chart saved to: `/Users/rabahbabaci/Developement/MTC-Transit_Access_Equity/visualizations/jobs30_bar_chart.png`



Bar Chart of Accessibility Statistics

This bar chart compares key accessibility statistics—mean, median, minimum, and maximum reachable jobs—between EPC and non-EPC tracts.

EPC tracts generally exhibit higher averages and medians, suggesting greater job accessibility within 30 minutes under the simplified distance-based model.

The figure (`jobs30_bar_chart.png`) is stored in the `visualizations/` folder for inclusion in the midterm report.

💡 7. Early Findings & Notes for Report

Use this section to summarize what you see on the map and in the data.

Example bullets you might write:

- EPC tracts cluster along Oakland and San Leandro corridors.
- Non-EPC tracts dominate eastern Alameda areas.
- Population density varies strongly between groups.

7. Early Findings & Notes

- **Spatial pattern (map)**
 - EPC tracts cluster along the West County corridor (Berkeley–Oakland–San Leandro–Hayward) near I-80/I-880 and major transit lines; fewer EPC tracts in the east (Castro Valley–Dublin–Pleasanton–Livermore).
 - EPC areas align with older, denser neighborhoods and job centers; non-EPC tracts dominate suburban/eastern parts of the county.
- **Accessibility pilot (jobs ≤ 30 min, 40 km/h proxy)**
 - Sampled EPC tracts averaged ~380 k reachable jobs, vs ~330 k for non-EPC (~+15 % gap).
 - *Interpretation:* EPC tracts are generally closer to employment clusters, even with a simple centroid-distance proxy.
- **Population context**
 - Many EPC tracts have moderate to high population; pairing high population with higher job accessibility suggests strong demand for transportation and services.
- **Data joins & keys**
 - Consistent 11-digit GEOID was crucial. We rebuilt GEOIDs via zero-padding (state = 2, county = 3, tract = 6) to merge TIGER tracts + EPC + ACS + LODES without mismatches.
 - LODES raw WAC required block→tract aggregation (`w_geocode` → first 11 chars; sum `C000`).
- **Map takeaways for planning**
 - EPC corridors overlap areas where EV charging + transit integration could serve many residents and jobs.
 - Non-EPC eastern tracts show lower accessibility—candidates for first/last-mile or targeted job access strategies.
- **Limitations (methods)**
 - Travel time proxy uses straight-line centroid distance and constant speed (no network or congestion effects).
 - LODES year = 2021 (post-COVID shifts possible).
 - EPC flag simplified (class labels). Threshold sensitivity could be tested.
- **Quality checks**
 - CRS harmonized to EPSG:4326 for storage; EPSG:26910 for distance.
 - Visual inspection shows clean tract coverage and EPC highlighting where expected (west county urban core).

Quantitative Results

Group	Tract Count	Mean Jobs ≤ 30 min	Median	Min	Max
EPC	97	419 217	421 795	321 849	482 405
Non-EPC	282	344 052	365 831	53 305	488 078

Interpretation: EPC tracts can reach $\approx 75\,000$ more jobs within 30 minutes than non-EPCs (~22 % higher accessibility), reflecting stronger links to employment along the western corridor (Berkeley–Oakland–San Leandro–Hayward).

Boxplot Interpretation

- EPC tracts show higher medians and tighter spread → consistently high accessibility.
- Non-EPC tracts show wider variance → mixed access and suburban isolation.

Key Takeaway EPC-designated tracts exhibit substantially higher job access within 30 minutes, highlighting transit and charging-equity needs in lower-access (non-EPC) areas.

Data sources: MTC Plan Bay Area 2050 (EPC), US Census TIGER/ACS 2022, LODS 2021 (LEHD).

Saved findings summary to `results/early_findings.md`



8. Next Steps (For Final Phase — Track B)

Goal: Generate time-aware accessibility metrics using R5 and GTFS/OSM.

Planned Tasks:

1. Download and unzip 511 GTFS feeds + OpenStreetMap extract.
2. Build `TransportNetwork` in `r5py`.
3. Compute travel-time matrices for AM/Mid/PM/Late periods.
4. Generate 15 / 30 / 45 min isochrones for each sample tract centroid.
5. Join results with LODS (job counts) and POI data (health + grocery).
6. Calculate accessibility gaps (EPC vs non-EPC).
7. Visualize with animated map (Plotly, Kepler.gl, or Pydeck).



9. Documentation & Data Dictionary

Column	Description	Example
GEOID	Census tract ID	06001400100
is_epc	1 if Equity Priority Community	0
population	Total population (ACS)	4295
geometry	Census tract boundary	POLYGON(...)

CRS: EPSG:4326 (WGS84)

Saved data dictionary to docs/data_dictionary.txt

✓ 10. Midterm Submission Checklist (Due Nov 10)

- ✓ Cleaned dataset → data/processed/alameda_final.geojson
- ✓ Interactive map → visualizations/epc_map.html
- ✓ Boxplot → visualizations/jobs30_boxplot.png
- ✓ Choropleth → visualizations/jobs30_map.png
- ✓ Bar chart → visualizations/jobs30_bar_chart.png
- ✓ Data dictionary → docs/data_dictionary.txt
- ✓ Findings summary → results/early_findings.md
- ✓ Short report (PDF, 4–5 pages) → results/report.pdf

🌐 11. Transition to GitHub (Track B Workflow)

1. Create a new private repo on GitHub (via classroom link or manual).
2. Add folders: notebooks/ , data/ , visualizations/ , docs/ , results/ .
3. Commit this notebook to notebooks/ .
4. Add README.md with project summary and data sources.
5. Include requirements.txt (or environment.yml).
6. Push to GitHub and enable GitHub Pages for hosting final site.
7. Each teammate works on a branch → merge via Pull Requests.

📖 12. References

- Metropolitan Transportation Commission (MTC) — Equity Priority Communities
- US Census Bureau — TIGER/Line Shapefiles, American Community Survey
- 511 SF Bay Open Data — GTFS Feeds
- OpenStreetMap — Network & POI Data
- Urban Institute — LODES Employment Data
- r5py Documentation — <https://r5py.readthedocs.io>