Deliverable 1

Transit Performance - Team C

Chandrahas, Eason, Manushi, Munir, Patrick

Introduction to project

- Public transit provides vital access to jobs, services, etc in cities like Boston
- Historically, some communities are underserved by public transit due to inequitable planning
- This analysis examines MBTA bus performance and census demographics in Boston
- Goals are to identify service disparities and see if they disproportionately affect marginalized groups
- Focus on Jan 2022 MBTA data and neighborhood-level census demographics in Boston
- Key questions:
    - What are the end-to-end travel times for different bus routes?
    - Are there noticeable disparities in on-time performance between routes?
    - What are the demographics of communities served by each route?
    - Do any service disparities appear to adversely impact disadvantaged groups?
- Uses data science techniques like APIs, data cleaning, EDA, mapping to answer questions
- Aims to provide insights into service inequities and guide investments to improve equity
- Public transit equity has implications for social justice, mobility, sustainability in diverse cities

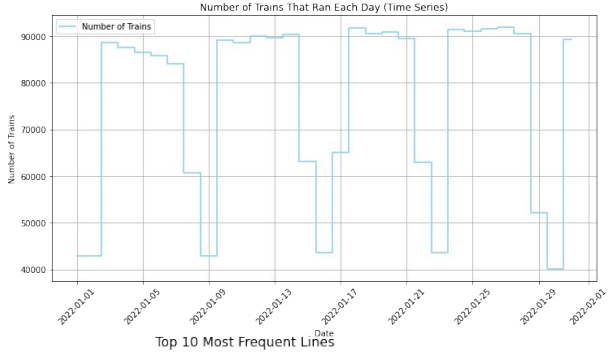
Data collection & processing

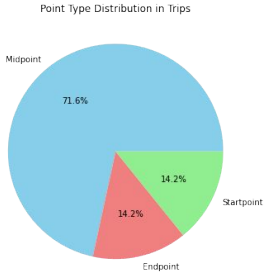
Some of the steps we took to create and clean our dataset:

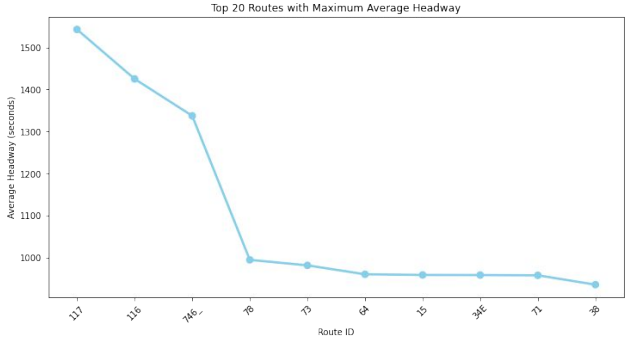
- Accessed MBTA API and extracted Jan 2022 bus data
- Filtered and aggregated raw data into consolidated dataset
- Calculated route travel times using scheduled and actual timestamps
- Computed on-time percentages by comparing scheduled vs actual

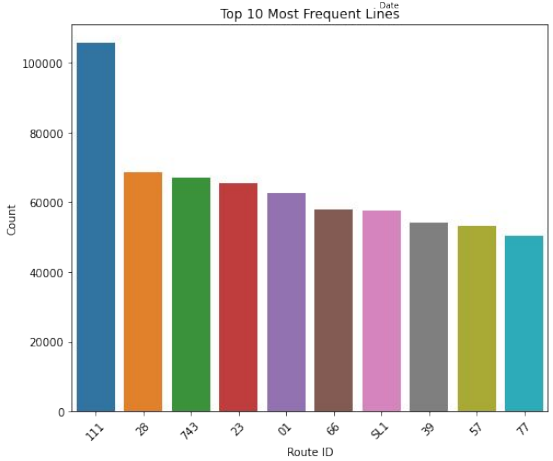
```python
import pandas as pd

def filter_data(df):
    filtered_df = df[df['point_type'].isin(['Startpoint', 'Endpoint'])]
    return filtered_df

def calculate_travel_times(df):
    grouped = df.groupby(['direction_id', 'half_trip_id'])
    travel_times = []

    for (direction, trip_id), group in grouped:
        startpoint = group[group['point_type'] == 'Startpoint']
        endpoint = group[group['point_type'] == 'Endpoint']

        startpoint_actual = startpoint['actual'].values[0]
        endpoint_actual = endpoint['actual'].values[0]
        service_date = group['service_date'].values[0]

        startpoint_actual = pd.to_datetime(startpoint_actual, format='%Y-%m-%d %H:%M:%S.%f')
        endpoint_actual = pd.to_datetime(endpoint_actual, format='%Y-%m-%d %H:%M:%S.%f')

        travel_time = endpoint_actual - startpoint_actual
        if travel_time.days == 0:
            hours, remainder = divmod(travel_time.seconds, 3600)
            minutes, seconds = divmod(remainder, 60)
            formatted_travel_time = f'{hours:02d}:{minutes:02d}:{seconds:02d}'

        route_id = group['route_id'].values[0]

        travel_times.append({
            'half_trip_id': trip_id,
            'service_date': service_date,
```

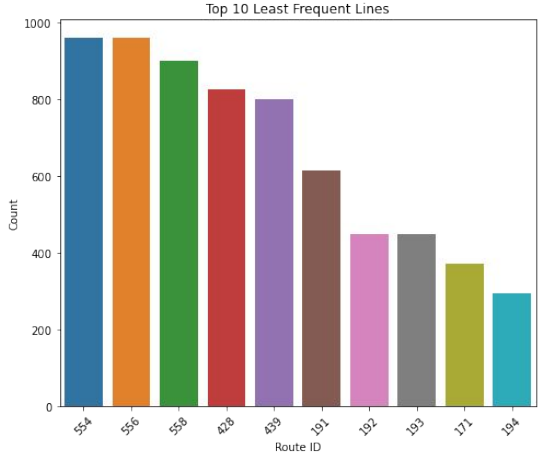
Slide on preliminary data analysis (graphs here)

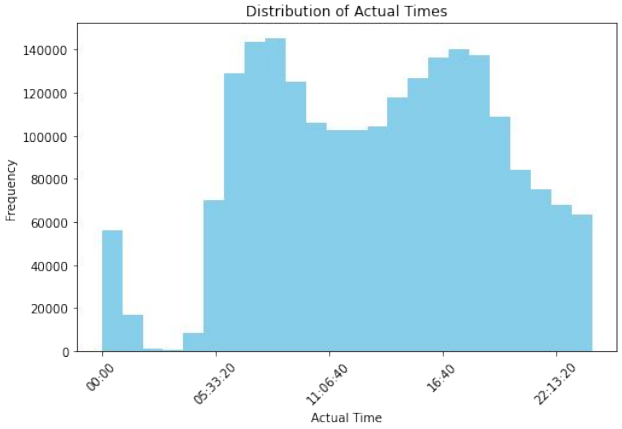




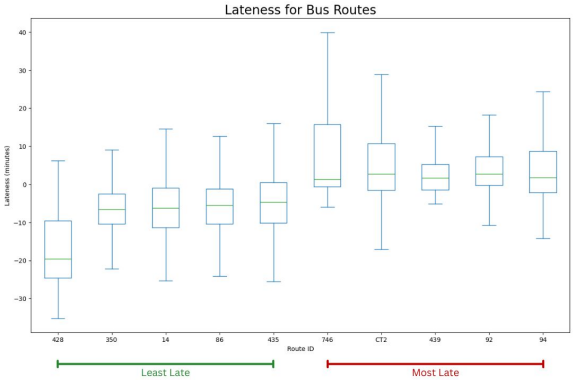


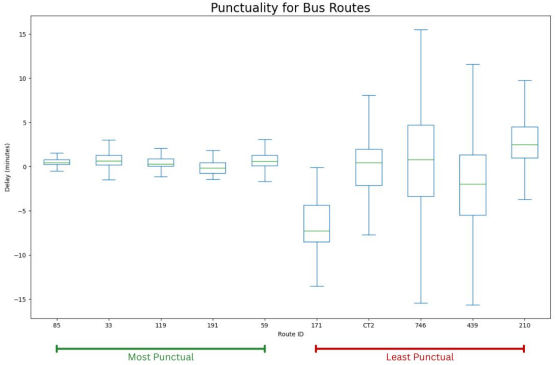


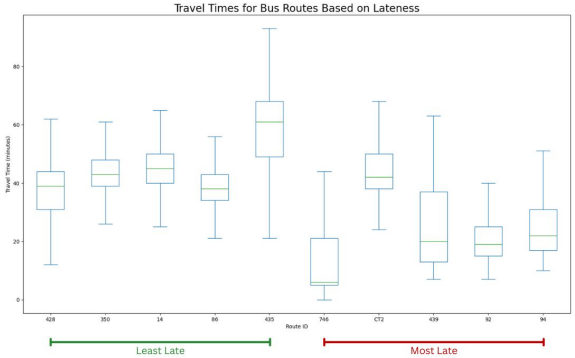


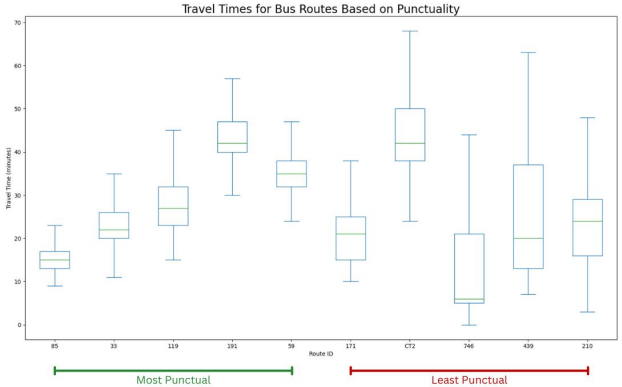


First two questions - analysis









What we've done, where we're going

- Calculated route travel times using MBTA scheduled and actual timestamps
- Computed on-time percentages by comparing scheduled vs. actual times
- Evaluated lateness and punctuality for different routes to identify disparities
- Preliminary findings show below average on-time service in some areas
- Next steps involve connecting MBTA data to census demographics factors