Lyft Bay Wheels

KyuRi Kim, Tianhao Wu

Project Overview

- Dock-based bikesharing systems
- Focused on Bay Area
- Main concern:
  - Bike using behaviors
  - Overused/Underused docks

Data Inputs

1. Lyft Bay Wheels Dataset:

1 Month data

- Start/End Time and Date
- Start/End Station ID
- Start/End Station Name
- Start/End Station Latitude
- Start/End Station Longitude
- Bike ID
- User Type

2. CensusGeoCode Package
3. CenPy Package

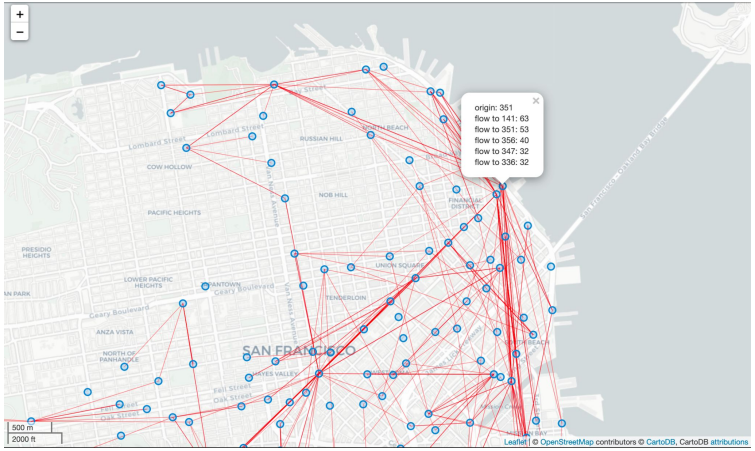
	ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual
0	9515818787165EA3	electric_bike	2021-09-22 13:30:57	2021-09-22 13:41:40	17th & Folsom Street Park (17th St at Folsom St)	SF-N23	Civic Center/UN Plaza BART Station (Market St ...	SF-I25	37.763686	-122.415521	37.780897	-122.412124	member
1	4857E21E7C02369C	electric_bike	2021-09-03 11:47:57	2021-09-03 11:54:02	The Embarcadero at Pier 38	SF-H30	4th St at 16th St	SF-M30	37.783198	-122.387994	37.767314	-122.390958	member
2	6AF0D6AC2DCF9B1B	classic_bike	2021-09-22 11:51:39	2021-09-22 12:06:54	27th St at MLK Jr Way	OK-I4	Telegraph Ave at 49th St	OK-D4	37.817015	-122.271761	37.835750	-122.262654	casual
3	B92F4F9D28E81783	docked_bike	2021-09-06 12:53:27	2021-09-06 12:59:23	The Embarcadero at Pier 38	SF-H30	The Embarcadero at Pier 38	SF-H30	37.782926	-122.387921	37.782926	-122.387921	casual
4	D5C84175F52EFBEF	classic_bike	2021-09-05 11:04:51	2021-09-05 11:09:00	The Embarcadero at Pier 38	SF-H30	Berry St at 4th St	SF-K29-1	37.782926	-122.387921	37.775880	-122.393170	member

Methods

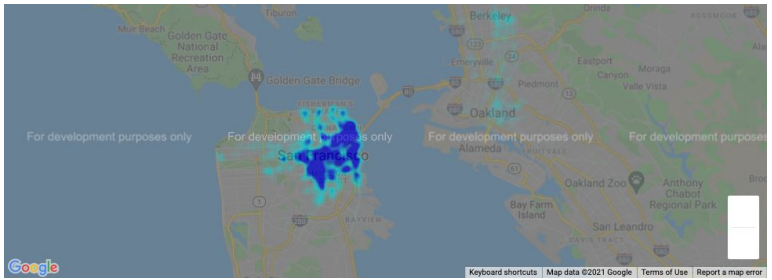
Used following methods to analyze the data:

- Flow DataFrame
- Travel Behavior Analysis
- Trip Distribution
- Clustering Census Data
- Network Analysis

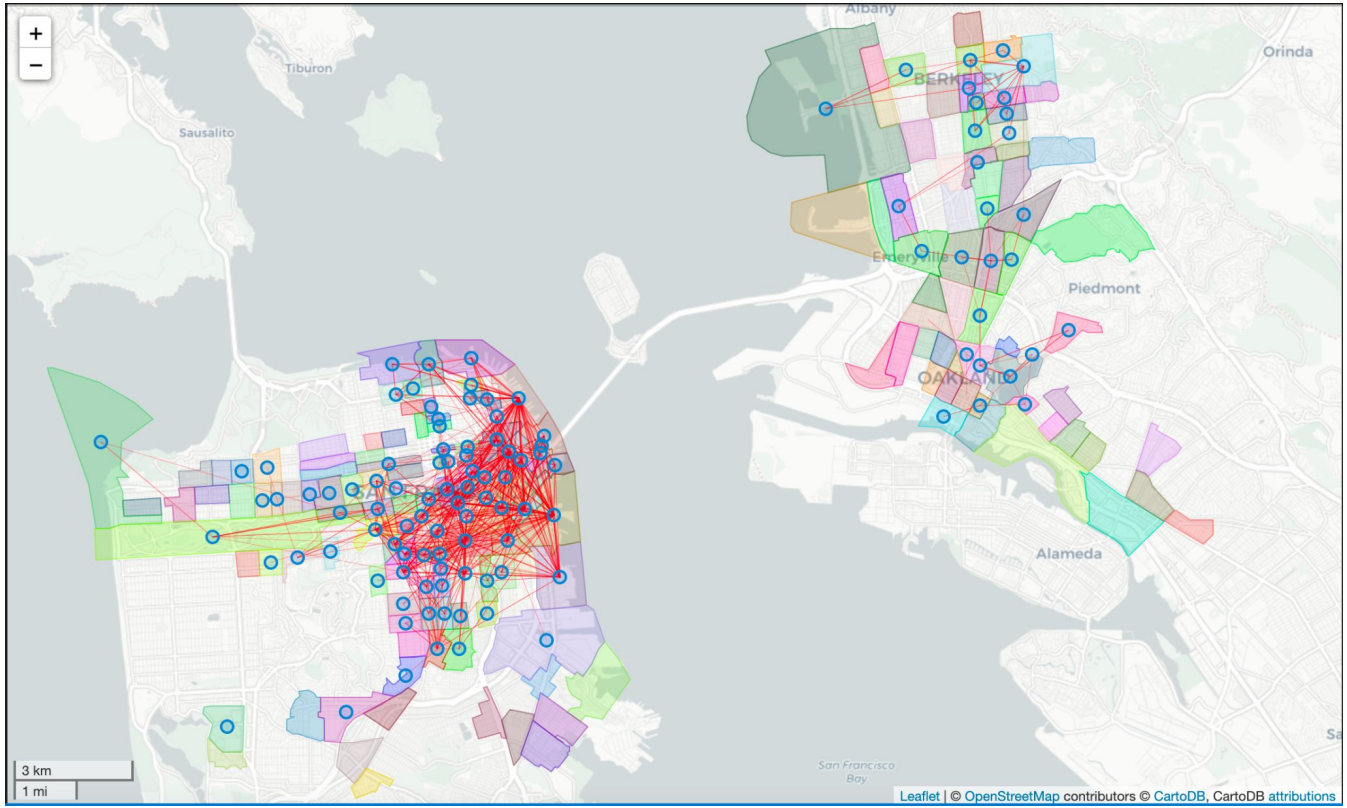
Data Cleaning & Integration



1. Dropped 65,699 NaN/Invalid rows
2. Left with 146,813 rows of trips
3. Find coordinate of station by averaging trip coordinates
4. Find geocode of stations given coordinates (CensusGeoCode)
5. Integrate stations into census tracts/geocode



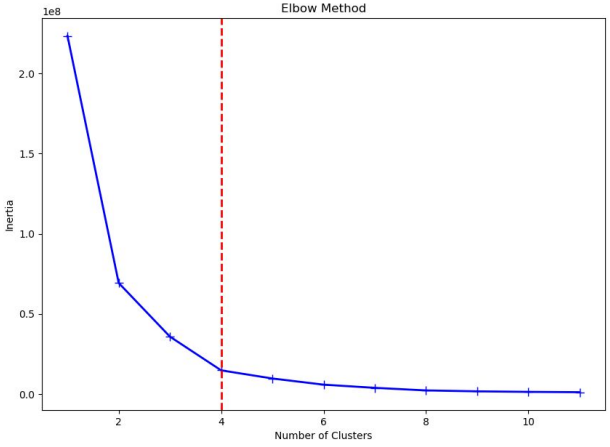
Trip Distribution

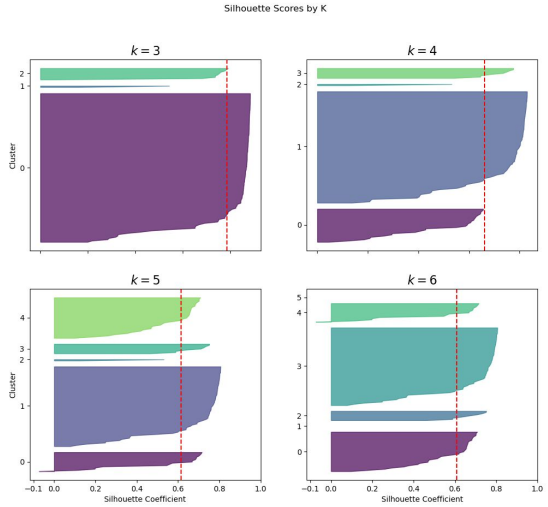


K-Means Clustering

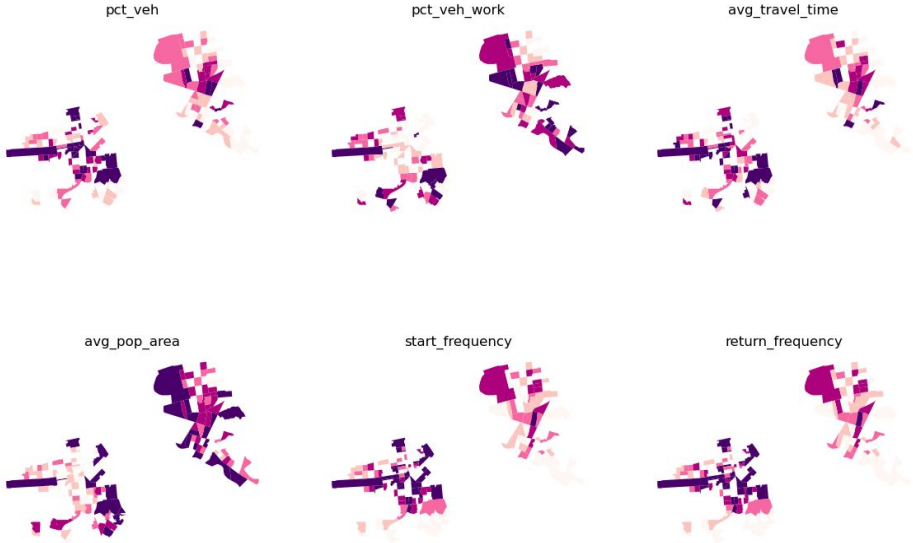
Data queried from Cenpy:	**Processed Variables:**
# Vehicles / Total population	Average # Vehicles owned
# Vehicles used in commuting / Total Population	Average # Vehicles used in commuting
Aggregated Travel Time / Total Population	Average Travel Time
Area sqm / Total Population	Land area per person
	Start Frequency
	End Frequency

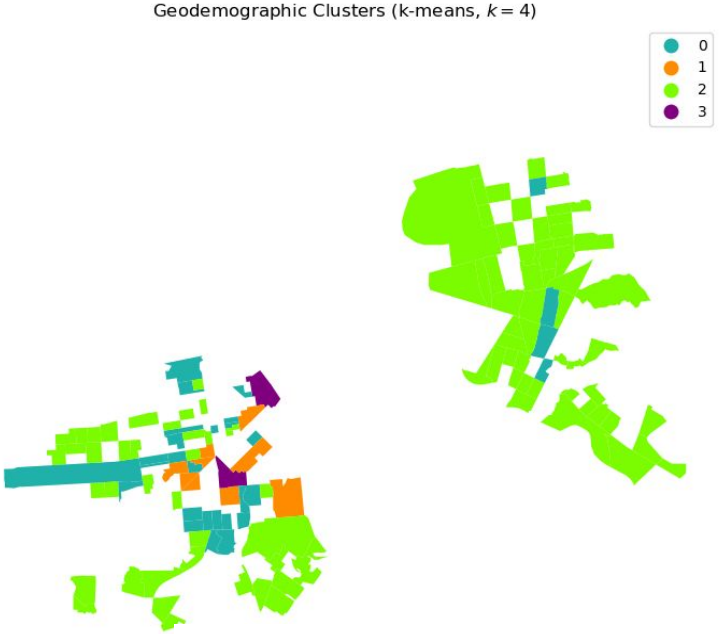
GEOID	geometry	total_pop	time_leave	time_travel	agg_travel_time	num_veh	num_veh_work	NAME
6081613600	POLYGON ((-13639358.830 4519608.220, -13639346...	6108.0	2557.0	2557.0	90860.0	2939.0	2100.0	Census Tract 6136, San Mateo County, California
6041113000	POLYGON ((-13661535.450 4581804.170, -13661410...	3371.0	1322.0	1322.0	48745.0	1504.0	1175.0	Census Tract 1130, Marin County, California

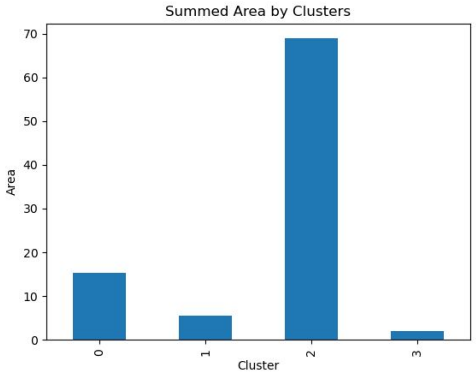




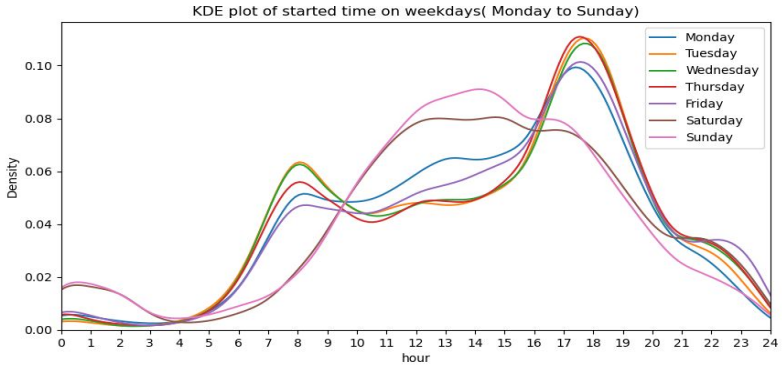
K-Means Clustering

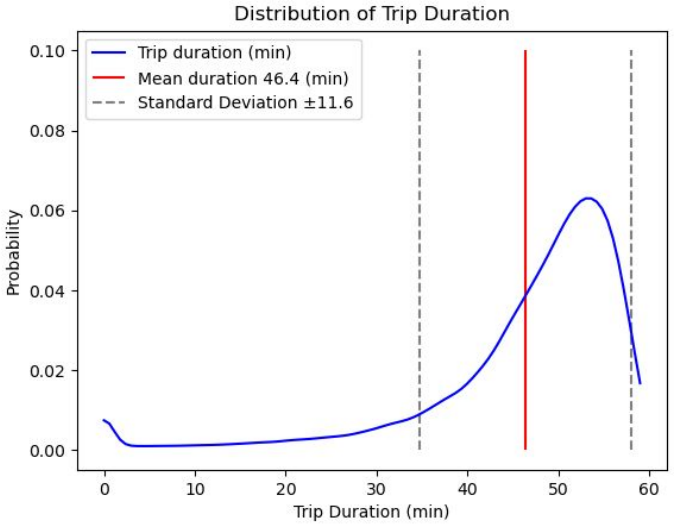




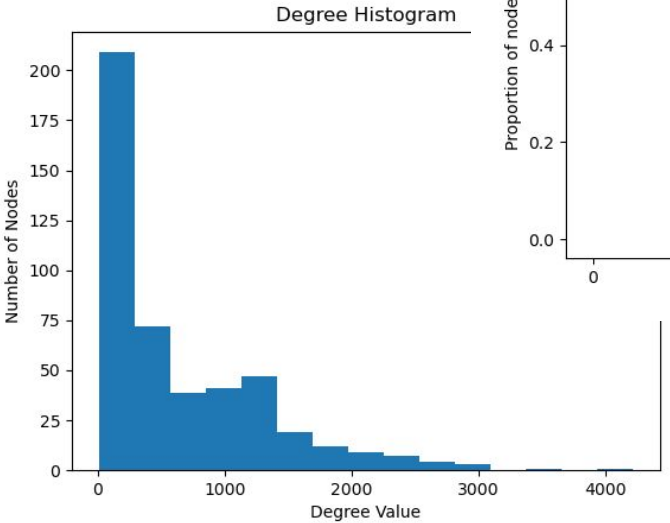


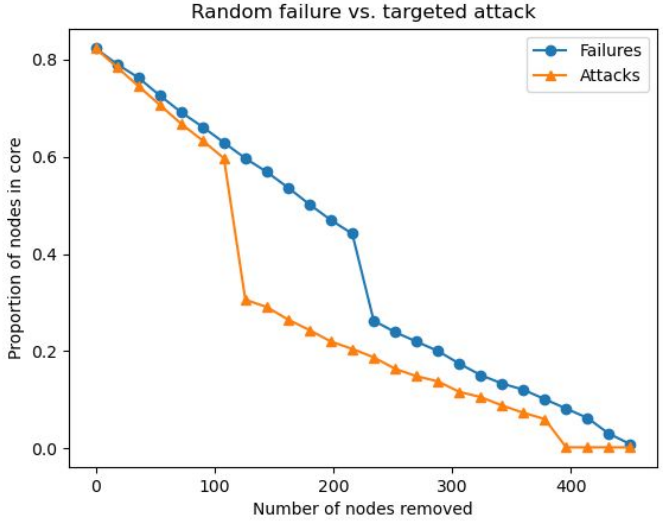
Travel Behaviors





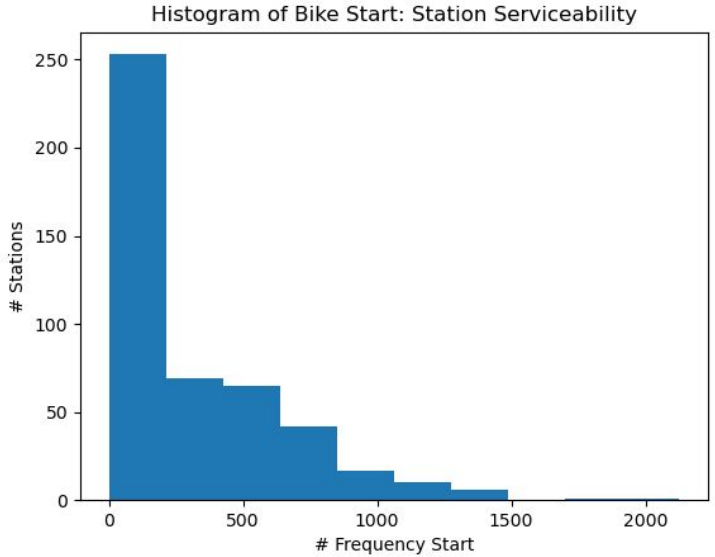
Network Analysis





- Nodes = Stations
- Edges = Trips
- G=MultiGraph

Station Serviceability



- Dock stations are mostly used lower than 250 times per month
- Two stations used more than 1500 times per month:
  - Market St at 10th St
  - Powell St BART Station (Market St at 4th St)

Questions?