Chicago Taxi Data

ACT 420 Final Project

Hasong Cho

Taxi Dataset

Source: City of Chicago Data Portal

Taxi ID	Trip Start Timestamp	Trip End Timestamp	Trip Seconds	Trip Miles	Pickup Community Area	Dropoff Community Area	Fare	Tips	Tolls	Extras	Trip Total	Payment Type
471b596f2a38426759f6f64c...	2016-02-26 19:30:00	2016-02-26 19:45:00	1,440.00	2.50	8.00	6.00	16.50	1.00	0.00	0.00	17.50	Credit Card
979ef85051ab2a0c1b287d2...	2016-02-26 21:30:00	2016-02-26 22:15:00	2,160.00	20.40	76.00	33.00	49.50	6.00	0.00	4.00	59.50	Credit Card
b83b4538c5567a8676f2cf3...	2016-02-13 16:15:00	2016-02-13 16:30:00	1,080.00	6.50	8.00	60.00	18.75	0.00	0.00	1.50	20.25	Cash
649d5ed9ff6292e7ab9392c...	2016-02-27 02:45:00	2016-02-27 02:45:00	480.00	1.60	6.00	3.00	7.75	1.00	0.00	1.00	9.75	Credit Card
29820405a25dab85dd7524...	2016-02-26 20:00:00	2016-02-26 20:15:00	960.00	2.30	8.00	28.00	11.25	0.00	0.00	0.00	11.25	Cash
...	...	...	...	...	...	...	...	...	...	...	...	...
0c6fcb78a456a7d83c720c7...	2022-12-31 23:45:00	2023-01-01 00:00:00	767.00	2.39	8.00	24.00	10.25	0.00	0.00	0.00	10.25	Cash
03b991b968c7f3e47e11691...	2022-12-31 23:45:00	2023-01-01 00:45:00	3,865.00	18.24	16.00	16.00	51.75	0.00	0.00	0.00	51.75	Cash
30e320f9c1c83229e437445...	2022-12-31 23:45:00	2023-01-01 00:15:00	974.00	1.69	32.00	8.00	9.50	0.00	0.00	1.50	11.00	Cash
fe89715757ed402a70b0f8b...	2022-12-31 23:45:00	2022-12-31 23:45:00	539.00	1.85	32.00	28.00	11.48	0.00	0.00	0.00	11.48	Mobile
3248d6ef1b3a20ada7e623e...	2022-12-31 23:45:00	2023-01-01 00:00:00	449.00	1.66	8.00	28.00	7.75	0.00	0.00	4.00	11.75	Cash

Year	Number of Rows	Columns
2023	5.5M	23
2022	6.38M	23
2021	3.95M	23
2020	3.89M	23
2019	16.5M	23
2018	20.7M	23
2017	25M	23
2016	31.8M	23

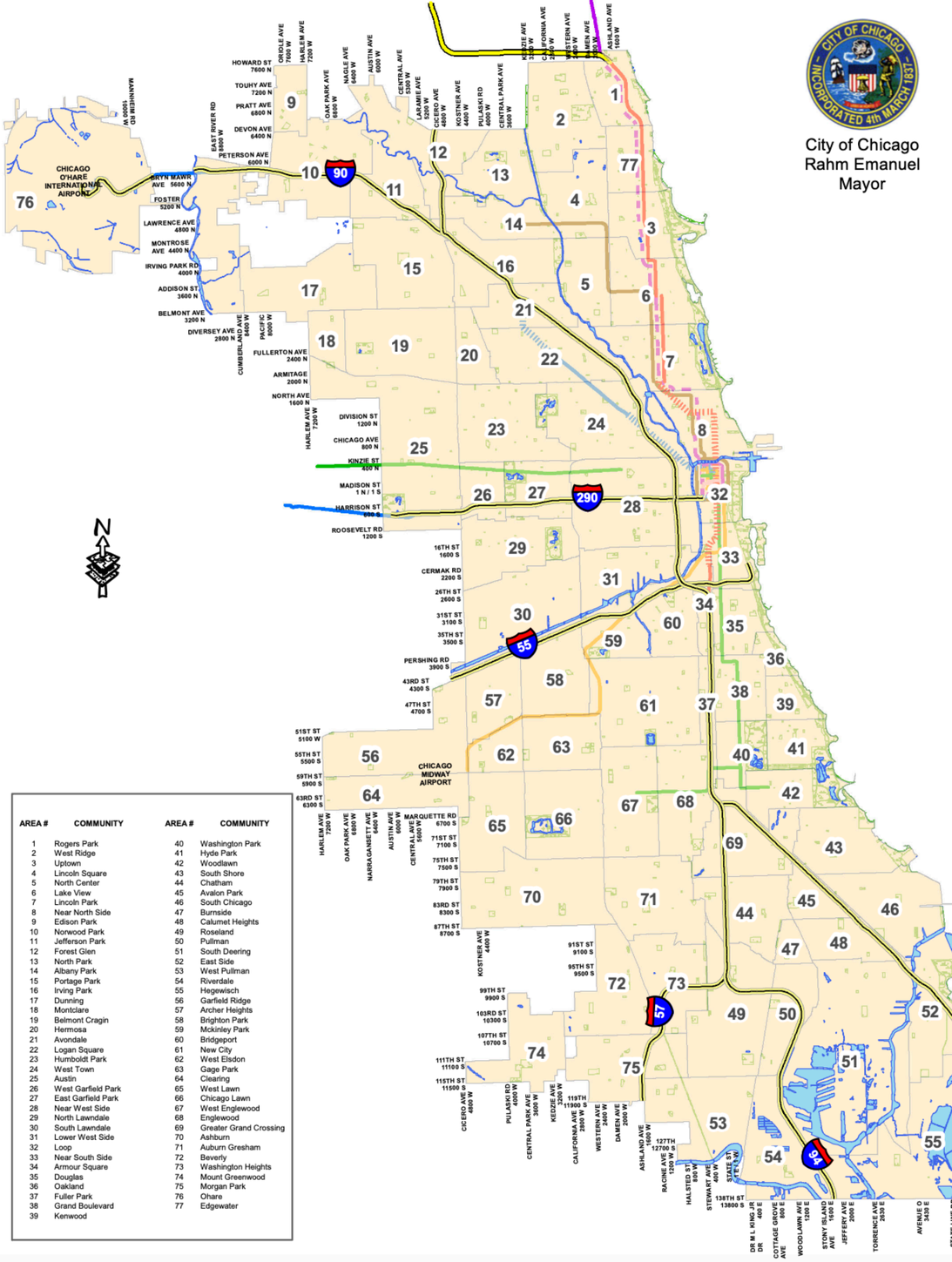
Category	Variables
Trip Basic Information	Trip Id, Trip Start Timestamp, Trip End Timestamp, Trip Seconds, Trip Miles
Trip Location Information	Pickup Census Tract, Dropoff Census Tract, Longitude / Latitude, Community Area
Taxi Information	Taxi Id, Company
Trip Price Information	Fare, Tips, Tolls, Extras, Trip Total, Payment Type

Community Area Dataset

Source: City of Chicago Data Portal

77 Community Areas

	AREA_NUMBE	COMMUNITY
0	35	DOUGLAS
1	36	OAKLAND
2	37	FULLER PARK
3	38	GRAND BOULEVARD
4	39	KENWOOD
...	...	...
72	74	MOUNT GREENWOOD
73	75	MORGAN PARK
74	76	OHARE
75	77	EDGEWATER
76	9	EDISON PARK



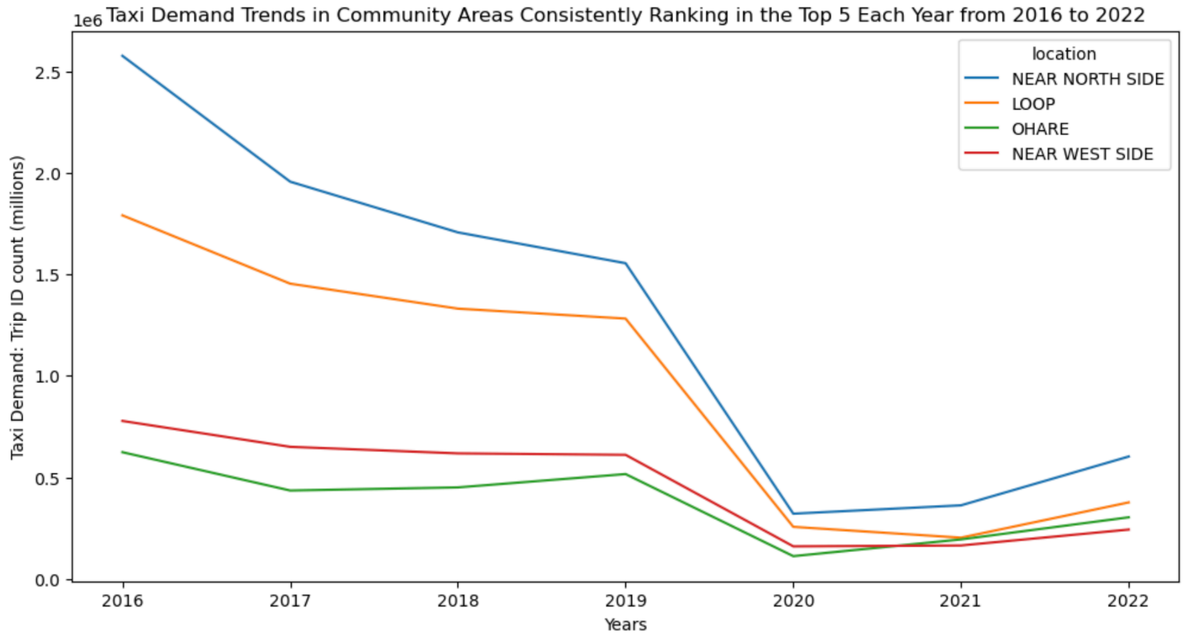
Question 1:

What pattern in taxi demand become evident when analyzing the pickup community areas between 2016 and 2022? Are there particular community areas experiencing notable increase in taxi demand?

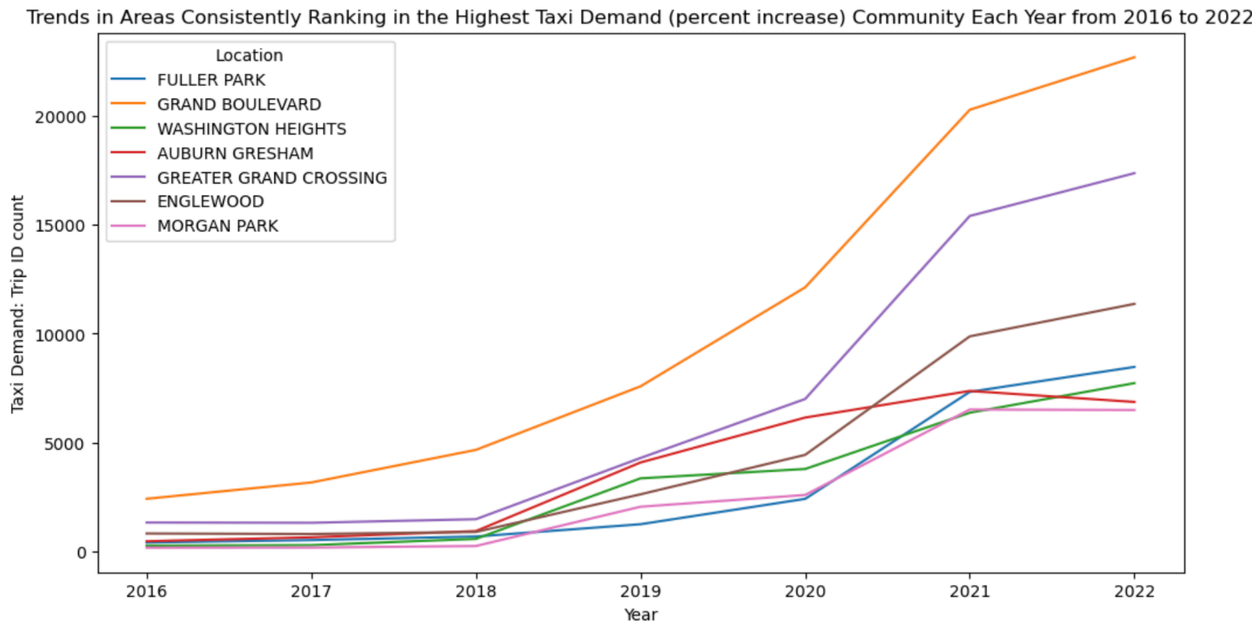
Implication:

- Increase in demand -> change in landscape -> urban development
- City of Chicago needs to be responsible for regulating the increase of taxi cabs usage in certain areas for efficient planning.

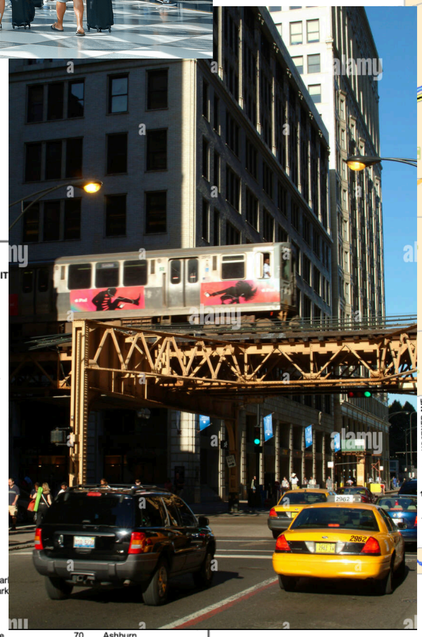
Definition 1: define taxi demand as number of trips, measured by Trip ID count. The more number of trips, the higher the demand.

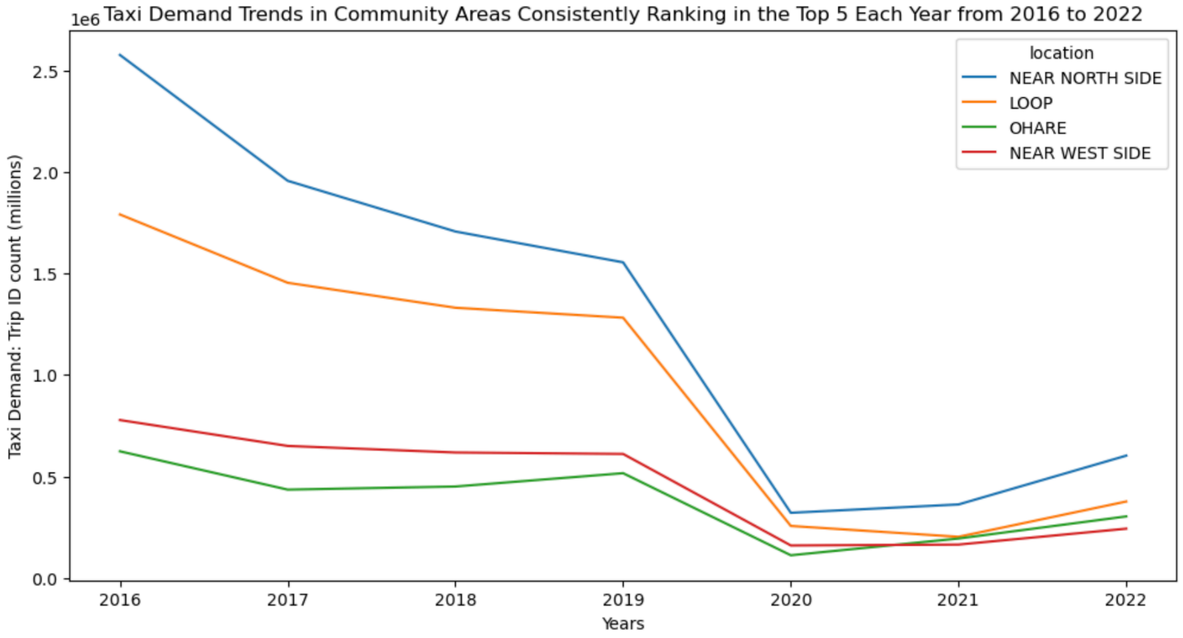


Definition 2: define taxi demand as percent change in number of trips. The higher the increase in percent change, the higher the demand.



Definition 1: define taxi demand as number of trips, measured by Trip ID count. The more number of trips, the higher the demand.

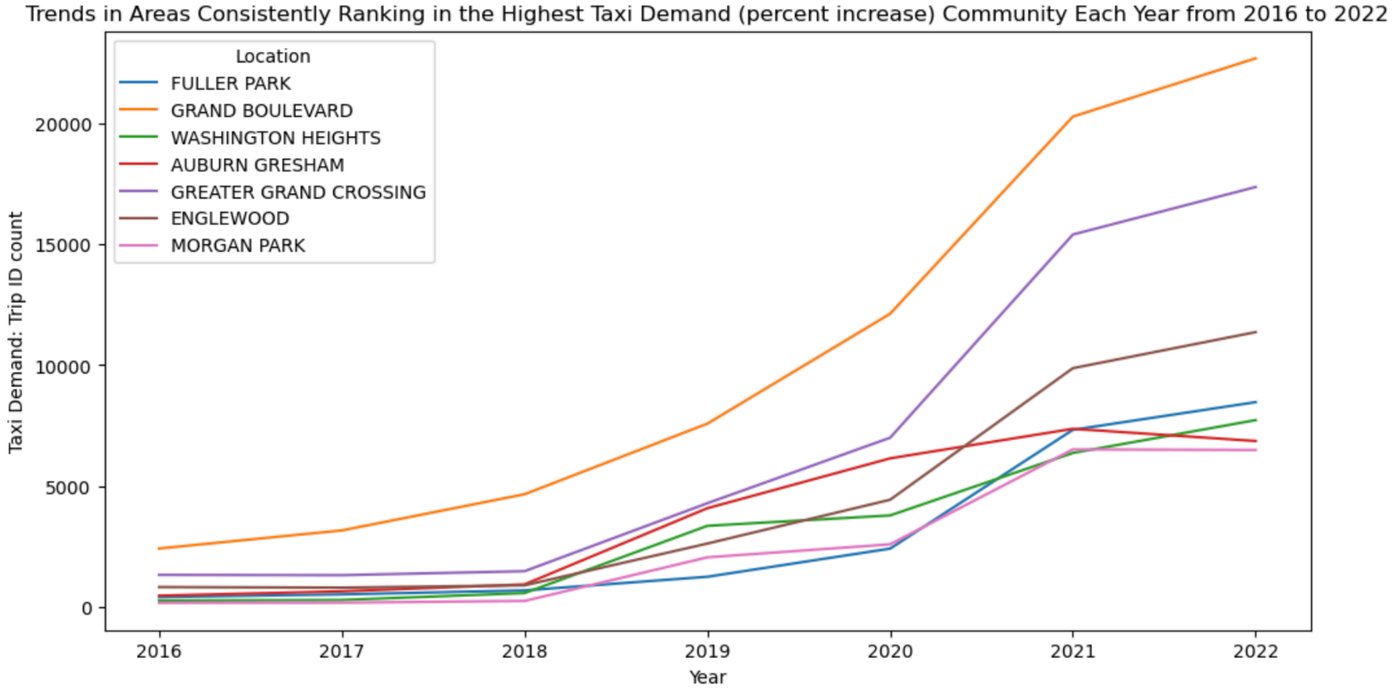




Definition 2: define taxi demand as percent change in number of trips. The higher the increase in percent change, the higher the demand.

City of Chicago
Rahm Emanuel
Mayor

AREA #	COMMUNITY	AREA #	COMMUNITY
1	Rogers Park	40	Washington Park
2	West Ridge	41	Hyde Park
3	Uptown	42	Woodlawn
4	Lincoln Square	43	South Shore
5	North Center	44	Chatham
6	Lake View	45	Avalon Park
7	Lincoln Park	46	South Chicago
8	Near North Side	47	Burnside
9	Edison Park	48	Calumet Heights
10	Norwood Park	49	Roseland
11	Jefferson Park	50	Pullman
12	Forest Glen	51	South Deering
13	North Park	52	East Side
14	Albany Park	53	West Pullman
15	Portage Park	54	Riverdale
16	Irving Park	55	Hegewisch
17	Dunning	56	Garfield Ridge
18	Montclare	57	Archer Heights
19	Belmont Cragin	58	Brighton Park
20	Hermosa	59	Mckinley Park
21	Avondale	60	Bridgeport
22	Logan Square	61	New City
23	Humboldt Park	62	West Elsdon
24	West Town	63	Gage Park
25	Austin	64	Clearing
26	West Garfield Park	65	West Lawn
27	East Garfield Park	66	Chicago Lawn
28	Near West Side	67	West Englewood
29	North Lawndale	68	Englewood
30	South Lawndale	69	Greater Grand Crossing
31	Lower West Side	70	Ashburn
32	Loop	71	Auburn Gresham
33	Near South Side	72	Beverly
34	Armour Square	73	Washington Heights
35	Douglas	74	Mount Greenwood
36	Oakland	75	Morgan Park
37	Fuller Park	76	Ohare
38	Grand Boulevard	77	Edgewater
39	Kenwood		



OLS: Taxi Demand Pattern from Monday to Sunday

OLS Regression Results

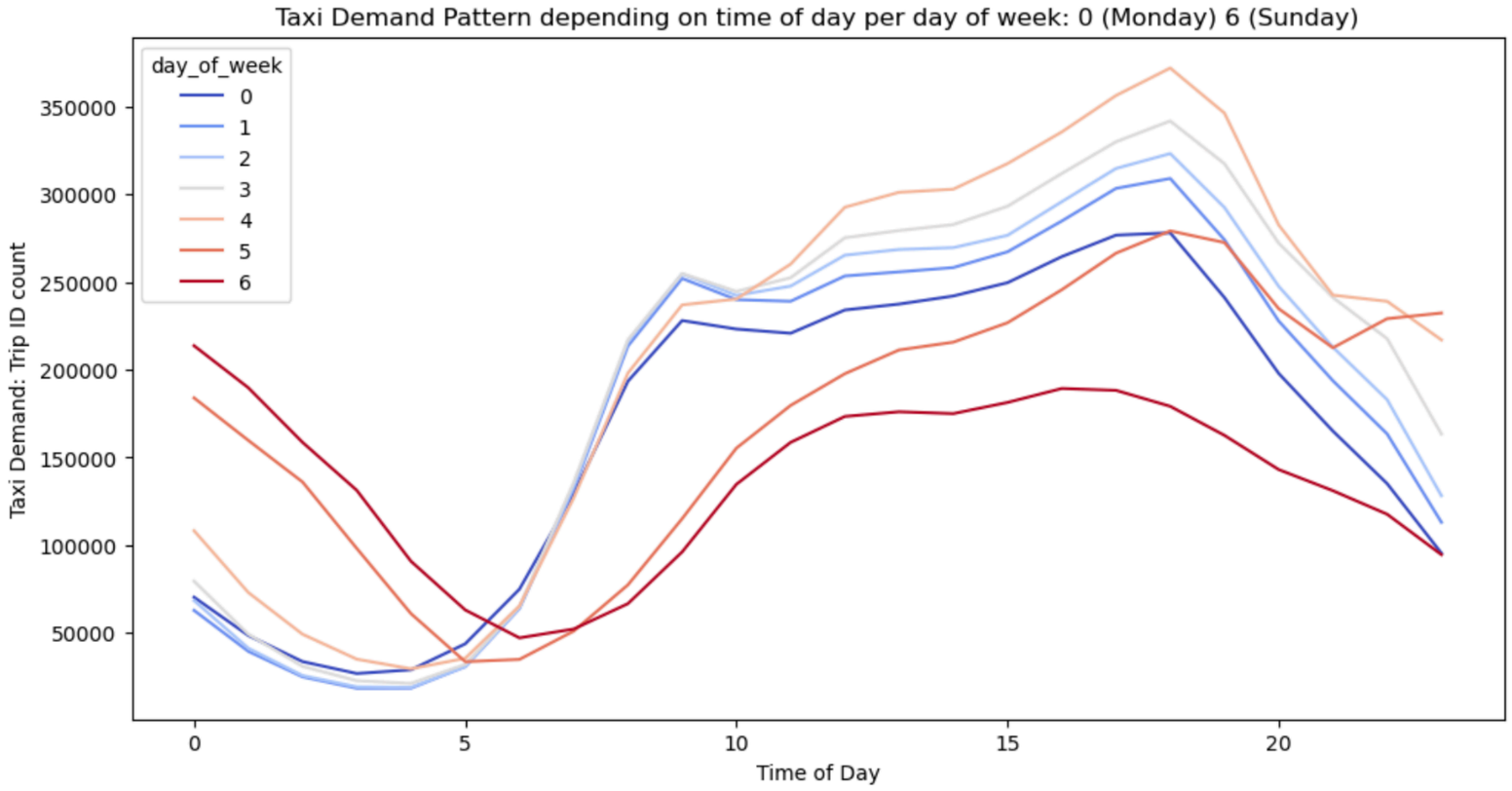
Dep. Variable:	y	**R-squared:**	0.559
Model:	OLS	**Adj. R-squared:**	0.559
Method:	Least Squares	**F-statistic:**	1.258e+07
Date:	Tue, 12 Dec 2023	**Prob (F-statistic):**	0.00
Time:	00:58:19	**Log-Likelihood:**	-2.2851e+08
No. Observations:	29835277	**AIC:**	4.570e+08
Df Residuals:	29835273	**BIC:**	4.570e+08
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	433.2711	0.308	1405.876	0.000	432.667	433.875
trip_miles	104.6036	0.017	6129.333	0.000	104.570	104.637
time_of_day	7.6179	0.016	469.184	0.000	7.586	7.650
day_of_week	-8.5445	0.049	-172.784	0.000	-8.641	-8.448

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.697
Model:	OLS	**Adj. R-squared:**	0.697
Method:	Least Squares	**F-statistic:**	2.283e+07
Date:	Tue, 12 Dec 2023	**Prob (F-statistic):**	0.00
Time:	01:20:05	**Log-Likelihood:**	-2.2292e+08
No. Observations:	29835277	**AIC:**	4.458e+08
Df Residuals:	29835273	**BIC:**	4.458e+08
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	240.4273	0.131	1836.851	0.000	240.171	240.684
trip_miles	-3.8918	0.033	-118.324	0.000	-3.956	-3.827
peak	224.3640	0.334	672.600	0.000	223.710	225.018
trip_total	43.6187	0.012	3659.617	0.000	43.595	43.642




Thank you