

Understanding the Dynamics of Social Encounters in NYC Public Spaces Using Human Mobility Data

Instructor: Takahiro Yabe, Graham Dove

Mentor: Vaidehi Raipat

Student: Ruixin Gan, Zheyang Chen, Yin Wang



Background

Background

Extreme Heat in NYC

Due to New York City's dense population and the corresponding rise in greenhouse gas emissions, the city is expected to experience more frequent, prolonged, and intense heat waves. These extreme temperatures pose an increasing threat to the health of residents. Prolonged exposure to high temperatures can lead to dehydration, heat exhaustion, heatstroke, and even fatalities.

Impact of Extreme Heat on Urban Parks

Extreme heat challenges park usability, equity, and infrastructure planning, affecting visitor behavior and thermal comfort.

Research Focus

Analyzing crowd distribution to understand how visitor behavior varies with temperature. Providing actionable insights for optimizing park design, resource allocation, and thermal comfort.

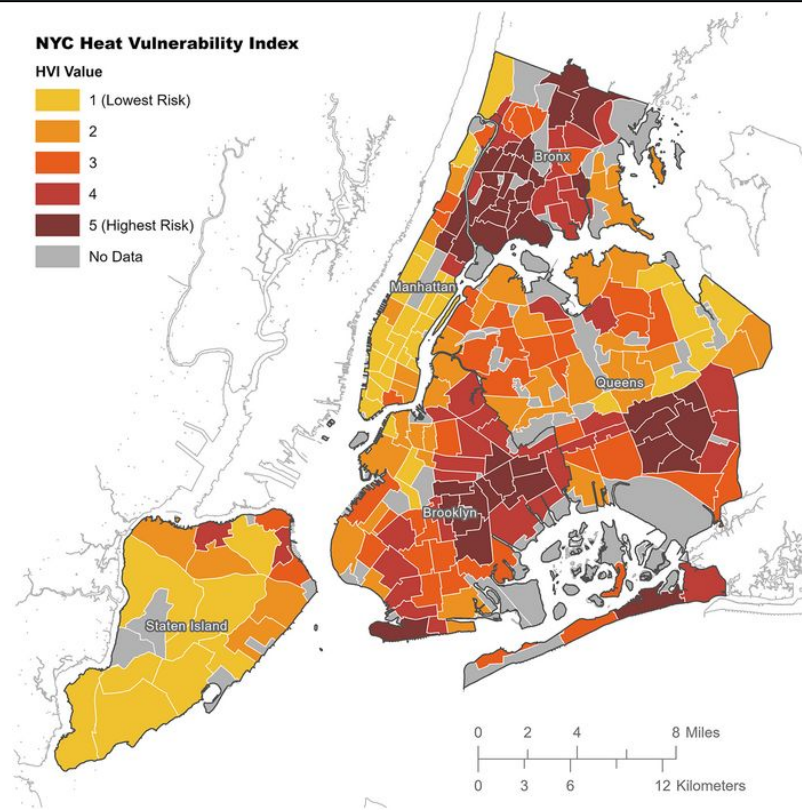


Research Question

Research Question

How do visitor behaviors, such as park usage patterns, time of visit, and distance traveled,vary between hot days, average-temperature days, and cooler days during the summer season?

- **Question:**How does park visitation (visitors count) differ between extremely hot and normal temperature days or hours during the summer?
- **Hypothesis:** Visitation decreases during hot hours but increases during cooler hours.
- **Question:** How is visitor density distributed across the park under different weather conditions (e.g., near shaded areas, open lawns, or water features)?
- **Hypothesis:** Visitors concentrate near shaded areas or water features during heat.



Methodology

Find Factor

Time Range: April 1 to September 29, 2024

Factor		
Shaded Area	Minimal	
	Partial	
	Fully	
Water Area	Water	
	Non-Water	
Date	Weekday	Include federal holidays
	Weekend	
Season	Summer	June-August
	Non-summer	April-May, September
Time	Morning	06:00-11:00
	Afternoon	12:00-17:00
	Evening	18:00-21:00
Heat	Extreme	Determined by heat index
	Hot	
	Average	
	Cool	
Weather	Clear	
	Cloudy	
	Rainy	
Wind	Windy	
	Non-Windy	
Event	Non-Event	Music Show (6.7-6.9/7.20) US Open (8.26-9.8) Boat Festival (8.3)
	Small Event	
	Large Event	

Heat Category

- **Extreme Hot:** Heat index $\geq 100^{\circ}\text{F}$ or $\geq 95^{\circ}\text{F}$ for two consecutive hours. (NYCEM)
- **Hot:** $100^{\circ}\text{F} > \text{Heat index} \geq 90^{\circ}\text{F}$
- **Average-Temperature:** $90^{\circ}\text{F} > \text{Heat index} \geq 80^{\circ}\text{F}$
- **Cooler:** Heat index $< 80^{\circ}\text{F}$

Cool	Average Temperature
13952h	2484h
Hot	Extreme Hot
528h	172h



**Flushing Meadows
Corona park POI**



Prospect Park POI



Kissena Park POI

POI List with Shaded Level and Water feature

Park	Name by Land Use	Type	Land Cover	Large Event	Small Event	Hour Visits Aggregated by Week	Water Area
Flushing Meadow Park	Water Front Zone_FM	Custom	Lakeside with trails	06-24 +220%,	05-01 +106% , 08-31 +122%	04-15 ~ 09-29	1
	Athletic Zone_FM	Custom	Open fields with trees	06-07 ~ 06-09 Music Show,		T	0
	Plaza Zone_FM	Custom	Open plazas with trees	06-07 ~ 06-09 Music Show, 08-26 ~ 09-08 US Open		T	1
	Recreation Zone_FM	Custom	Open meadow with trees	07-20 Music Show, 08-03 Boat Festival	05-26 +117%,	T	1
Prospect Park	Entrance Zone_P	Custom	Entrance/Forested with trails	04-28 Half Marathon,	06-22 +112%	T	0
	Trail Zone_P	Custom	Forested with trails	04-28 Half Marathon		T	1
	Athletic Zone_P	Custom	Open fields with trees		06-15 +109%	T	0
	Villa Green Zone_P	Custom	Playgrounds/Forested with trails	07-30 +410%	06-15 +104%, 06-22 +112%,	T	0
	Lakeside Zone_P	Custom	Lakeside with trails			T	1
	Meadows Zone_P	Custom	Open meadow with trees	04-28 Half Marathon,	06-22 +175%	T	1
Kissena Corridor Park	Athletic Zone_K	Custom	Open Field with Tree			T	0
	Nature and Trail Zone_K	Custom	Lakeside with Forest and Trail			04-08 ~ 09-22	1



Regression Result

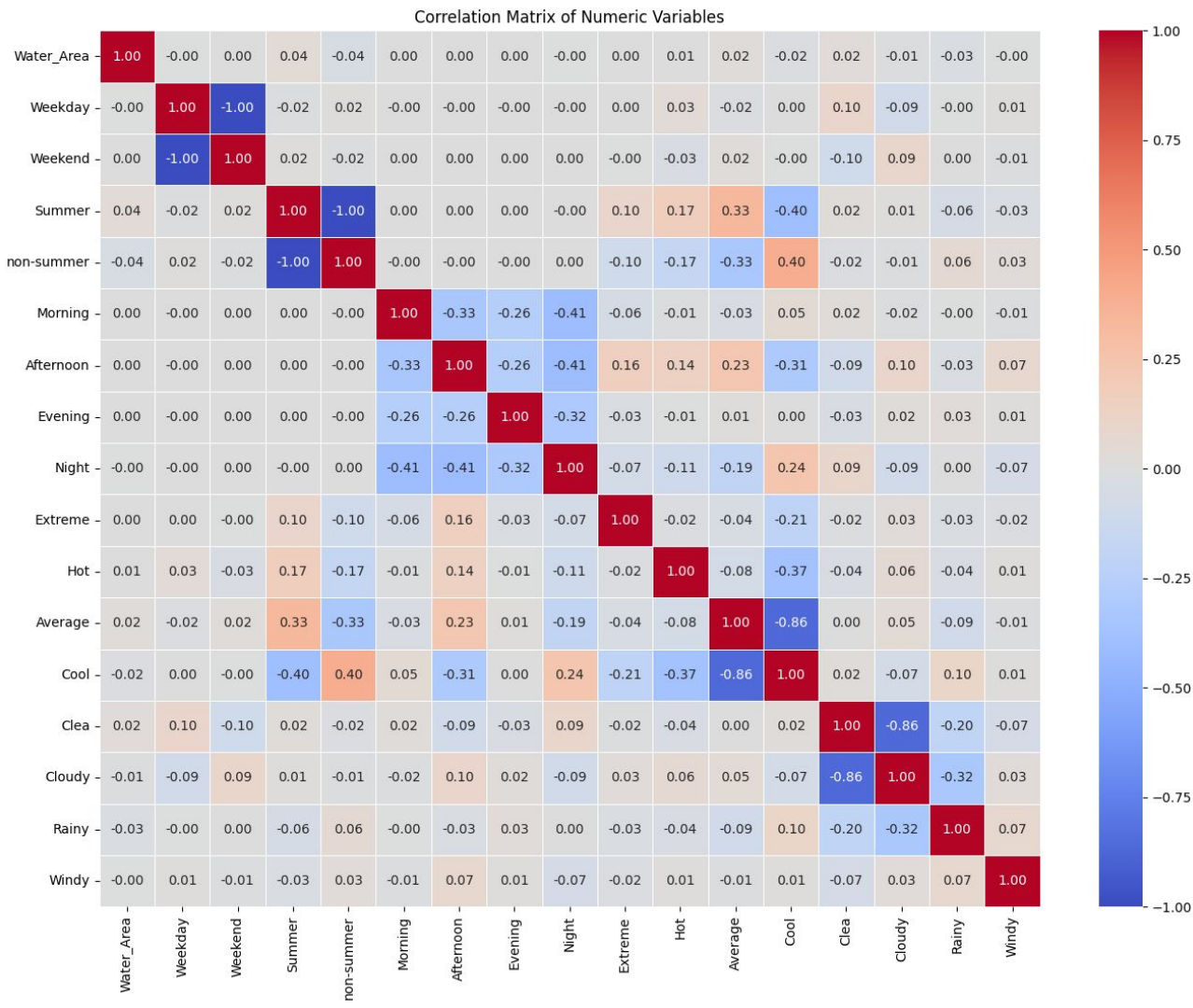
Correlation

Kissena Park

Kissena Park VIF

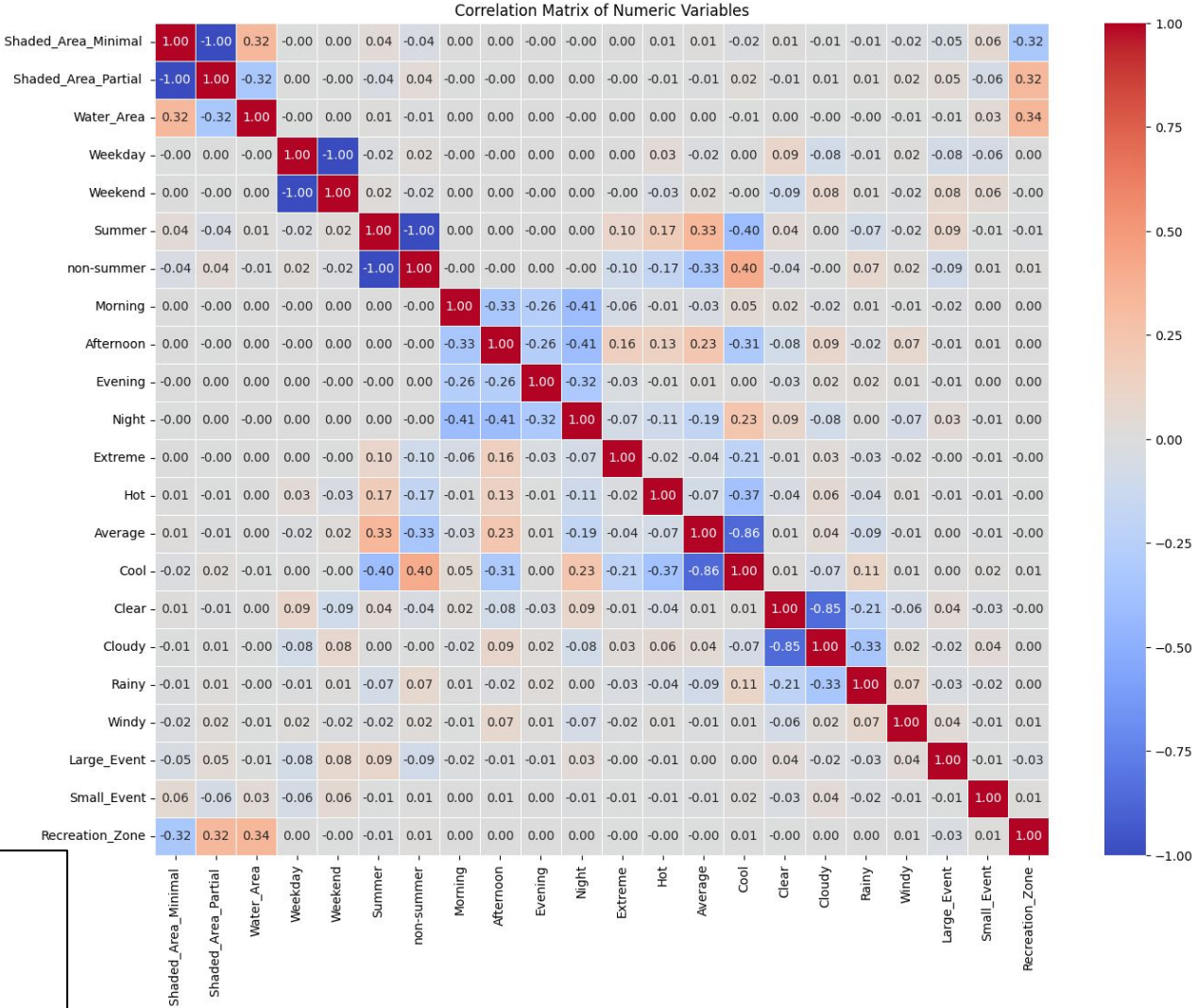
	feature	VIF
0	Weekend	1.461335
1	Water_Area	1.890397
2	Summer	2.225647
3	Evening	1.473279
4	Morning	1.686629
5	Afternoon	1.835735
6	Extreme	1.124708
7	Average	2.237727
8	Cool	4.453650
9	Rainy	1.216735
10	Cloudy	2.605077
11	Windy	1.042021

Kissena Park:
Merge Water Area and Shaded
Area Features



Correlation Flushing Park

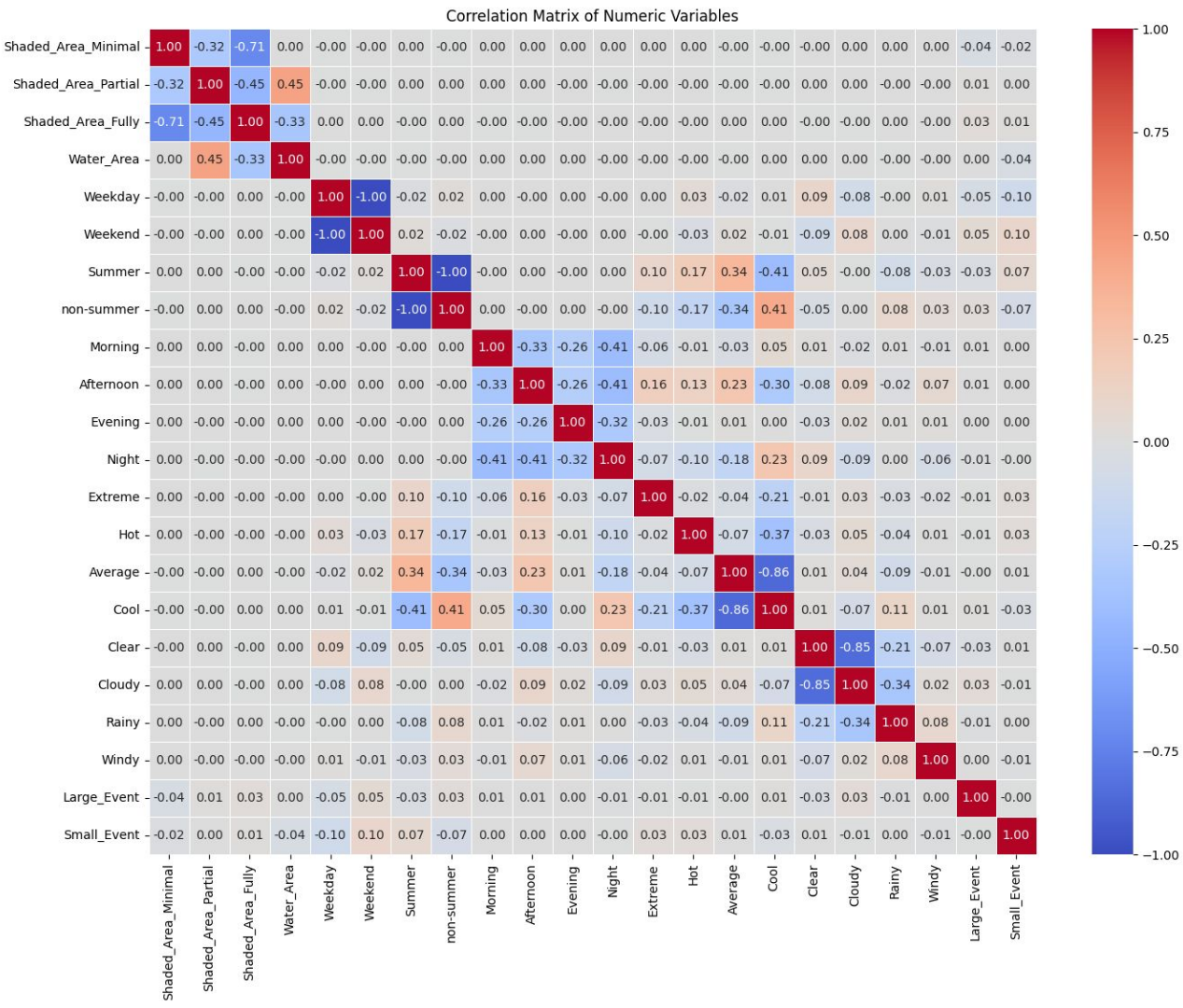
	feature	VIF
0	Shaded_Area_Minimal	1.942098
1	Water_Area	4.597724
2	Weekend	1.432558
3	Summer	2.231146
4	Morning	1.533182
5	Afternoon	1.809710
6	Evening	1.382066
7	Extreme	1.070447
8	Hot	1.135609
9	Average	1.498943
10	Cloudy	2.316230
11	Rainy	1.174039
12	Windy	1.041930
13	Large_Event	1.041537
14	Small_Event	1.011614
15	Recreation_Zone	2.008958



Flushing Park:
Add Recreation Zone POI Feature
(binary)

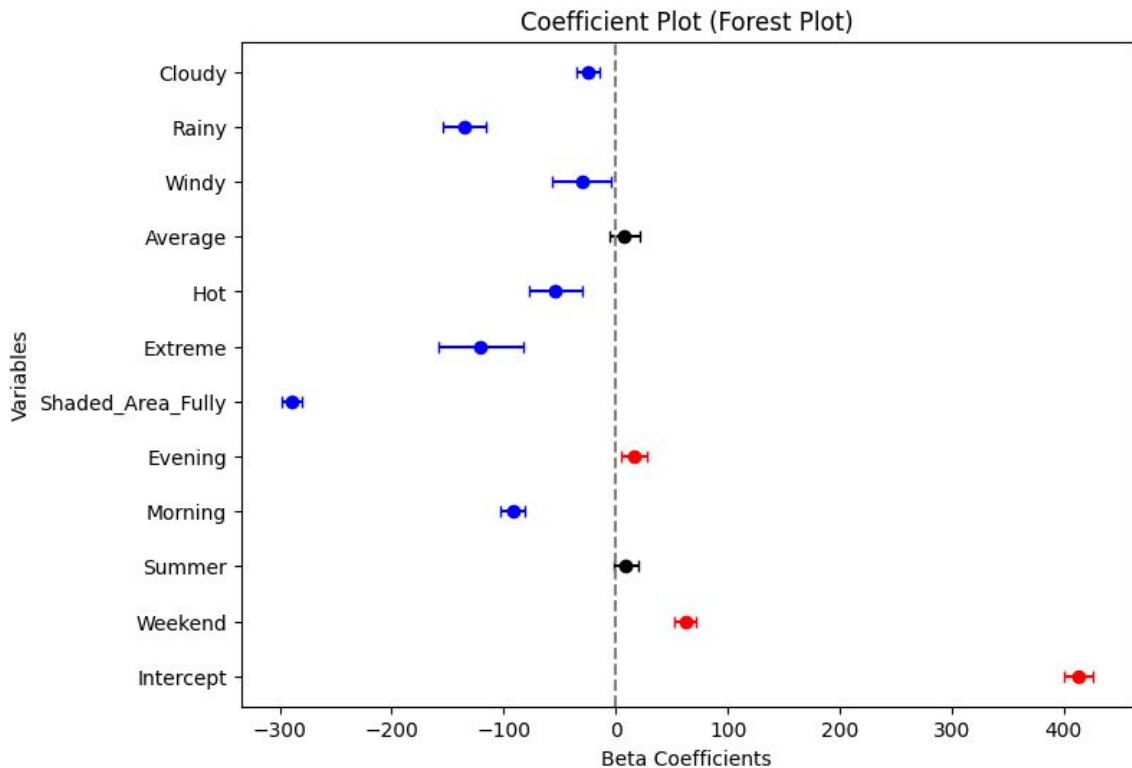
Correlation Prospect Park

	feature	VIF
0	Weekend	1.455780
1	Shaded_Area_Partial	1.693665
2	Shaded_Area_Fully	2.302187
3	Water_Area	2.424755
4	Summer	1.870509
5	Evening	1.407132
6	Morning	1.607892
7	Afternoon	1.656295
8	Extreme	1.071172
9	Hot	1.168311
10	Cool	3.763221
11	Rainy	1.240134
12	Cloudy	2.598959
13	Windy	1.042371
14	Large_Event	1.008779
15	Small_Event	1.022607



Kissena Park

Y = Visits



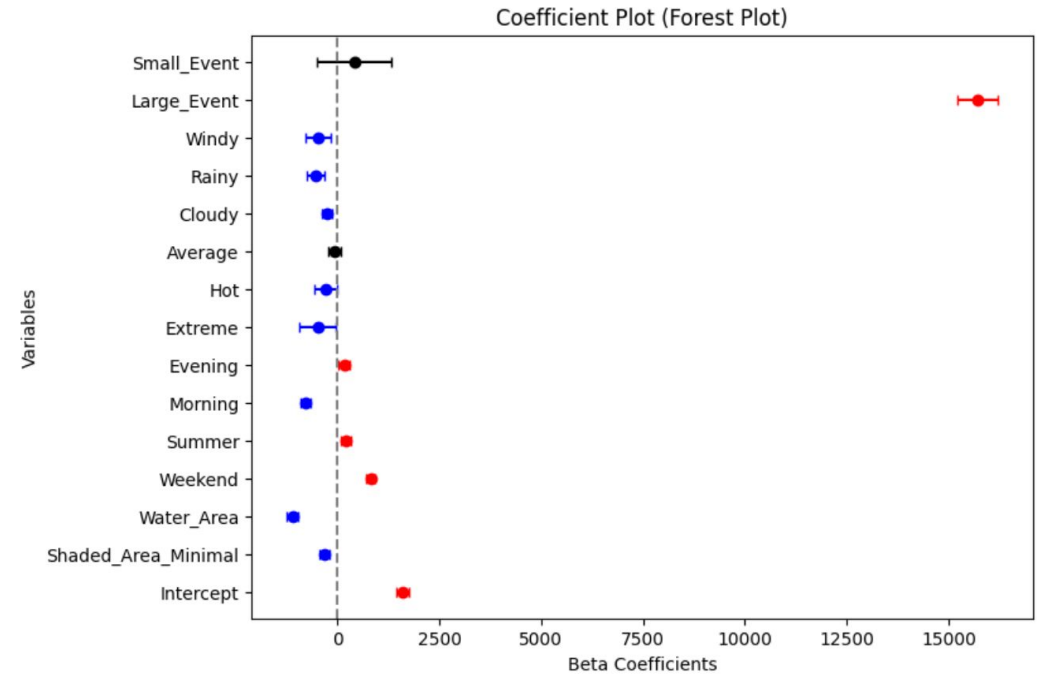
OLS Regression Results						
Dep. Variable:	Visits	R-squared:	0.452			
Model:	OLS	Adj. R-squared:	0.451			
Method:	Least Squares	F-statistic:	419.3			
Date:	Mon, 31 Mar 2025	Prob (F-statistic):	0.00			
Time:	20:30:44	Log-Likelihood:	-36807.			
No. Observations:	5600	AIC:	7.364e+04			
Df Residuals:	5588	BIC:	7.372e+04			
Df Model:	11					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	412.7392	6.397	64.520	0.000	400.199	425.280
Weekend	62.1802	5.063	12.281	0.000	52.254	72.106
Summer	9.2264	5.329	1.731	0.083	-1.221	19.674
Morning	-91.7825	5.621	-16.327	0.000	-102.803	-80.762
Evening	16.2103	6.157	2.633	0.008	4.140	28.281
Shaded_Area_Fully	-289.1490	4.641	-62.304	0.000	-298.247	-280.051
Extreme	-120.6213	19.525	-6.178	0.000	-158.897	-82.345
Hot	-53.7868	11.946	-4.503	0.000	-77.205	-30.369
Average	7.8254	6.750	1.159	0.246	-5.407	21.057
Windy	-30.4440	13.450	-2.263	0.024	-56.812	-4.076
Rainy	-134.6651	9.813	-13.723	0.000	-153.902	-115.428
Cloudy	-24.9198	5.123	-4.864	0.000	-34.963	-14.877
Omnibus:	1352.541	Durbin-Watson:	0.490			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	5041.590			
Skew:	1.166	Prob(JB):	0.00			
Kurtosis:	7.021	Cond. No.	13.1			

Percentage Table

Variable	Coefficient	Percentage	Interpretation
Intercept	412.74	-	Baseline prediction (Weekday, Non-Summer, Afternoon, Shaded_Area_Minimal, Clear & Cool Weather, Non-Windy).
Weekend	+62.18	+15.1%	Weekend visits are 15.1% higher than weekday baseline.
Morning	-91.78	-22.2%	Morning visits are 22.2% lower than afternoon baseline.
Evening	+16.21	+3.9%	Evening visits are 3.9% higher than afternoon baseline.
Shaded_Area_Fully	-289.15	-70.1%	Fully-shaded areas see 70.1% fewer visits than non-shaded areas.
Extreme	-120.62	-29.2%	Extreme weather reduces visits by 29.2% vs. moderate weather.
Hot	-53.79	-13.0%	Hot temperatures reduce visits by 13.0% vs. moderate temperatures.
Windy	-30.44	-7.4%	Windy conditions reduce visits by 7.4% vs. non-windy baseline.
Rainy	-134.67	-32.6%	Rainy days reduce visits by 32.6% vs. clear weather baseline.
Cloudy	-24.92	-6.0%	Cloudy weather reduces visits by 6.0% vs. clear weather baseline.

Flushing Park

Y = Visits

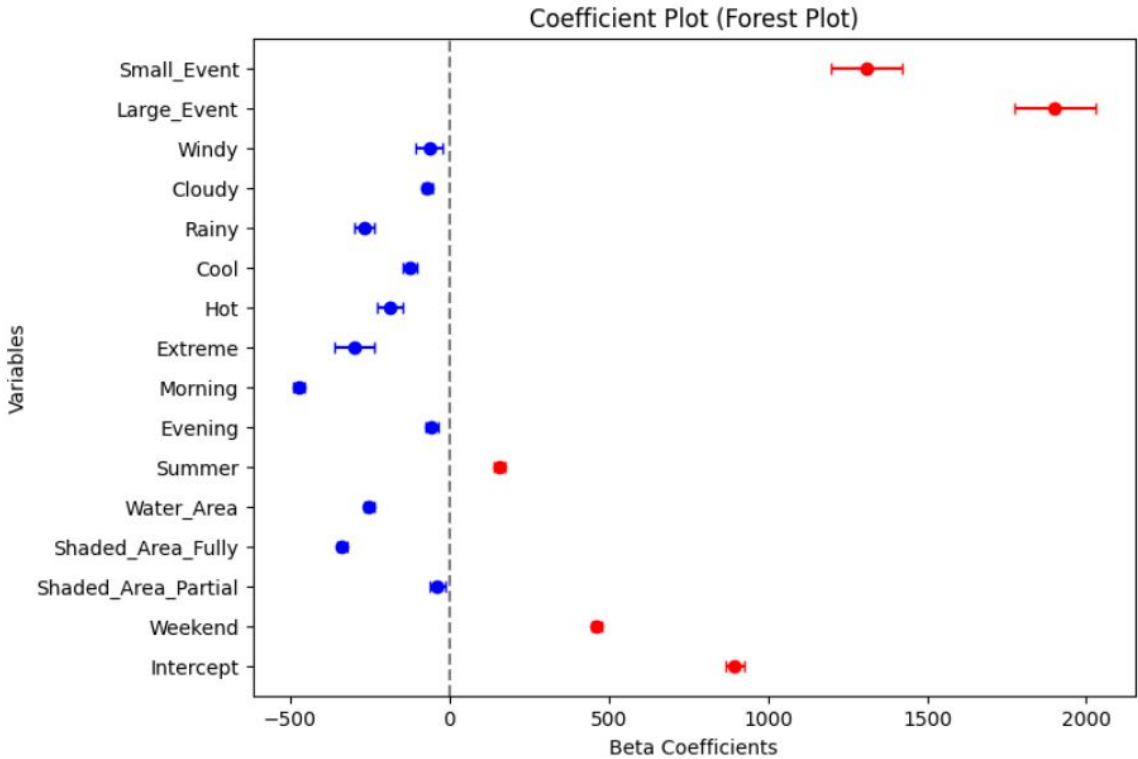


OLS Regression Results						
Dep. Variable:	Visits	R-squared:	0.309			
Model:	OLS	Adj. R-squared:	0.308			
Method:	Least Squares	F-statistic:	364.3			
Date:	Mon, 31 Mar 2025	Prob (F-statistic):	0.00			
Time:	19:43:51	Log-Likelihood:	-1.0727e+05			
No. Observations:	11424	AIC:	2.146e+05			
Df Residuals:	11409	BIC:	2.147e+05			
Df Model:	14					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	1597.7903	84.371	18.938	0.000	1432.409	1763.171
Shaded_Area_Minimal	-317.6069	67.704	-4.691	0.000	-450.318	-184.896
Water_Area	-1102.1972	65.753	-16.763	0.000	-1231.085	-973.310
Weekend	816.2640	59.621	13.691	0.000	699.396	933.132
Summer	221.0166	62.981	3.509	0.000	97.564	344.469
Morning	-777.8152	65.722	-11.835	0.000	-906.641	-648.989
Evening	174.6365	72.005	2.425	0.015	33.495	315.778
Extreme	-480.1169	230.764	-2.081	0.037	-932.454	-27.779
Hot	-277.0081	141.246	-1.961	0.050	-553.875	-0.141
Average	-77.6926	79.807	-0.974	0.330	-234.128	78.743
Cloudy	-254.2291	60.272	-4.218	0.000	-372.373	-136.085
Rainy	-527.3739	112.086	-4.705	0.000	-747.083	-307.665
Windy	-466.3650	161.481	-2.888	0.004	-782.896	-149.834
Large_Event	1.571e+04	252.514	62.223	0.000	1.52e+04	1.62e+04
Small_Event	419.7690	466.600	0.900	0.368	-494.847	1334.385
Omnibus:	23178.145	Durbin-Watson:	0.201			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	88986370.873			
Skew:	16.775	Prob(JB):	0.00			
Kurtosis:	434.069	Cond. No.	28.5			

Variable	Coefficient	Percentage	Interpretation
Intercept	1597.79	-	Baseline prediction (Weekday, Non-Summer, Afternoon, Shaded_Area_Partial, Non-Water_Area, clear and cool Weather, No Events).
Shaded_Area_Minimal	-317.61	-19.9%	Minimal-shade areas see 19.9% fewer visits than non-minimal shaded areas.
Water_Area	-1102.20	-69.0%	Waterfront areas have 69.0% fewer visits than non-water areas.
Weekend	+816.26	+51.1%	Weekend visits are 51.1% higher than weekday baseline.
Summer	+221.02	+13.8%	Summer visits are 13.8% higher than non-summer baseline.
Morning	-777.82	-48.7%	Morning visits are 48.7% lower than afternoon baseline.
Evening	+174.64	+10.9%	Evening visits are 10.9% higher than afternoon baseline.
Extreme	-480.12	-30.1%	Extreme weather reduces visits by 30.1% vs. moderate weather.
Hot	-277.01	-17.3%	Hot temperatures reduce visits by 17.3% vs. moderate temperatures.
Cloudy	-254.23	-15.9%	Cloudy weather reduces visits by 15.9% vs. clear weather baseline.
Rainy	-527.37	-33.0%	Rainy days reduce visits by 33.0% vs. clear weather baseline.
Windy	-466.37	-29.2%	Windy conditions reduce visits by 29.2% vs. non-windy baseline.
Large_Event	+15,710.00	+983.1%	Large events increase visits by 983.1%.

Prospect Park

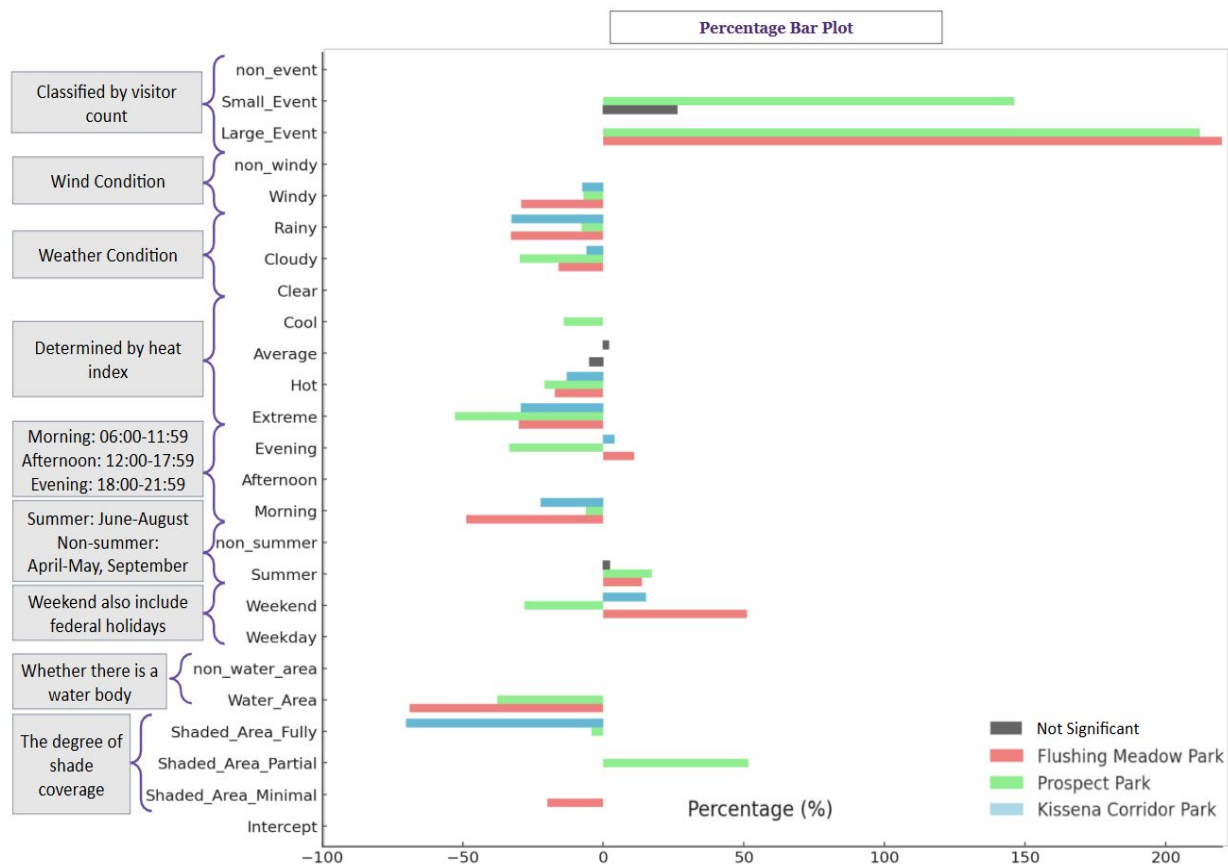
Y = Visits



OLS Regression Results

Dep. Variable:	Visits	R-squared:	0.395			
Model:	OLS	Adj. R-squared:	0.394			
Method:	Least Squares	F-statistic:	758.4			
Date:	Mon, 31 Mar 2025	Prob (F-statistic):	0.00			
Time:	19:47:20	Log-Likelihood:	-1.3361e+05			
No. Observations:	17472	AIC:	2.672e+05			
Df Residuals:	17456	BIC:	2.674e+05			
Df Model:	15					
Covariance Type: nonrobust						
	coef	std err	t	P> t	[0.025	0.975]
Intercept	895.7053	15.371	58.273	0.000	865.577	925.834
Weekend	462.2307	8.443	54.744	0.000	445.681	478.781
Shaded_Area_Partial	-38.4742	12.311	-3.125	0.002	-62.605	-14.343
Shaded_Area_Fully	-337.4379	8.706	-38.761	0.000	-354.502	-320.374
Water_Area	-252.0247	8.708	-28.942	0.000	-269.093	-234.956
Summer	155.3138	8.887	17.476	0.000	137.894	172.734
Evening	-55.9876	10.184	-5.498	0.000	-75.949	-36.026
Morning	-472.8818	9.291	-50.897	0.000	-491.093	-454.671
Extreme	-298.8062	32.978	-9.061	0.000	-363.447	-234.166
Hot	-187.2004	20.606	-9.085	0.000	-227.590	-146.811
Cool	-126.3951	11.382	-11.105	0.000	-148.704	-104.086
Rainy	-266.9462	15.701	-17.002	0.000	-297.722	-236.171
Cloudy	-69.3784	8.566	-8.099	0.000	-86.169	-52.588
Windy	-63.1910	22.268	-2.838	0.005	-106.839	-19.543
Large_Event	1900.7957	64.720	29.369	0.000	1773.937	2027.654
Small_Event	1309.8562	57.734	22.688	0.000	1196.692	1423.020
Omnibus:	13519.179	Durbin-Watson:	0.280			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	475630.096			
Skew:	3.420	Prob(JB):	0.00			
Kurtosis:	27.628	Cond. No.	29.9			

Variable	Coefficient	Percentage	Interpretation
Intercept	895.71	-	Baseline prediction (Weekday, Non-Summer, Afternoon, Shaded_Minimal Areas, Non-Water Areas, cool and clear Weather, No Events).
Weekend	+462.23	+51.6%	Weekend visits are 51.6% higher than weekday baseline.
Shaded_Area_Partial	-38.47	-4.3%	Partially-shaded areas see 4.3% fewer visits than Shaded_Minimal .
Shaded_Area_Fully	-337.44	-37.7%	Fully-shaded areas see 37.7% fewer visits than Shaded_Minimal .
Water_Area	-252.02	-28.1%	Waterfront areas have 28.1% fewer visits than non-water areas.
Summer	+155.31	+17.3%	Summer visits are 17.3% higher than non-summer baseline.
Evening	-55.99	-6.3%	Evening visits are 6.3% lower than afternoon baseline.
Morning	-472.88	-52.8%	Morning visits are 52.8% lower than afternoon baseline.
Extreme	-298.81	-33.4%	Extreme weather reduces visits by 33.4% vs. cool weather.
Hot	-187.20	-20.9%	Hot temperatures reduce visits by 20.9% vs. cool temperatures.
Cool	-126.40	-14.1%	Cool temperatures reduce visits by 14.1% vs. cool temperatures.
Rainy	-266.95	-29.8%	Rainy days reduce visits by 29.8% vs. clear weather baseline.
Cloudy	-69.38	-7.7%	Cloudy weather reduces visits by 7.7% vs. clear weather baseline.
Windy	-63.19	-7.1%	Windy conditions reduce visits by 7.1% vs. non-windy baseline.
Large_Event	+1900.80	+212.2%	Large events increase visits by 212.2%.
Small_Event	+1309.86	+146.2%	Small events increase visits by 146.2%.



Dashboard

Park Visitor Number Forecast Dashboard

Park:

- ☒ Flushing Meadow Park
- ☐ Prospect Park
- ☐ Kissena Corridor Park

Feature:

Season

- ☒ Summer
- ☐ non_summer

Temperature

- ☐ Extreme
- ☒ Hot
- ☐ Average
- ☐ Cool

Shaded

- ☐ Shaded_Area_Minimal
- ☒ Shaded_Area_Partial

Week

- ☒ Weekend
- ☐ Weekday

Weather

- ☒ Clear
- ☐ Cloudy
- ☐ Rainy

Event

- ☐ Large_Event
- ☐ Small_Event
- ☒ non_event

Time

- ☒ Morning
- ☐ Afternoon
- ☐ Evening

Wind Condition

- ☐ Windy
- ☒ non_windy

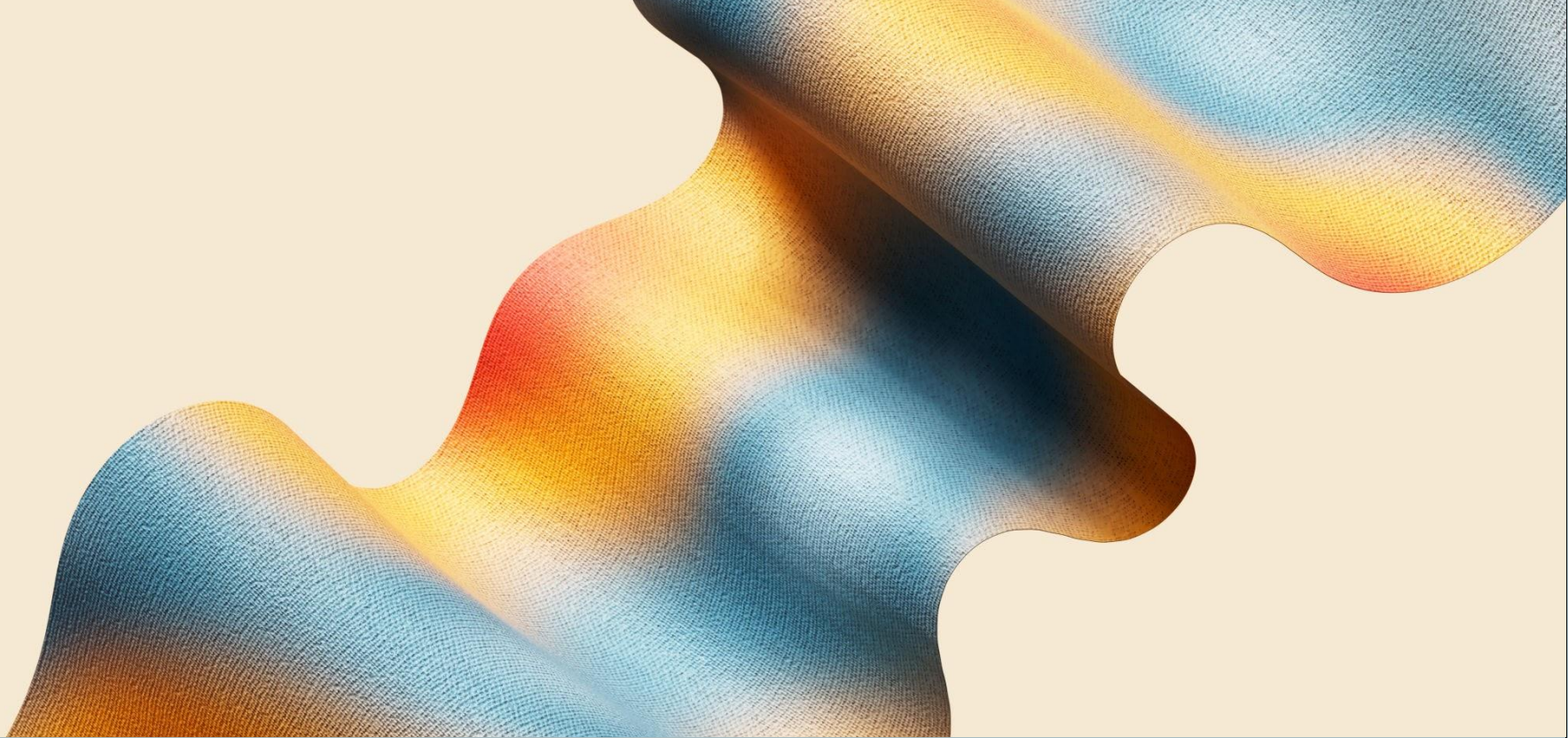
Water

- ☒ Water_Area
- ☐ non_water_area

Submit

Result

- Park:** Flushing Meadow Park
- Intercept Condition:** ['non_summer', 'Weekday', 'Afternoon', 'Cool', 'Clear', 'non_windy', 'Shaded_Area_Partial', 'non_event', 'Water_Area']
- Intercept Coefficient:** 495.59
- Feature Selected:** ['Summer', 'Weekend', 'Morning', 'Hot', 'Clear', 'non_windy', 'Shaded_Area_Partial', 'non_event', 'Water_Area']
- Visitor Number:** 478



Limitation & Next Step

Limitation

- **POI Constraints:**

Water bodies and dense shade areas limit usable space, lowering visits. Platform restrictions led to incomplete park coverage.

- **Inconsistent Baselines:**

Different reference categories across parks limit cross-site comparison.

- **Limited Explanatory Power:**

Moderate R^2 (0.31–0.40) suggests missing factors like transit access and land use.

Next Step

- **Demographic Pattern:**

Analysis of how visitor demographic characteristics (e.g., age distribution, family group composition) vary between hotter and cooler days.

- **Refine POI Zones:**

Group points of interest so each zone has similar visitor travel area.

- **Add New Factor:**

Bring in extra inputs like air quality and visitor spill-over before and after holidays to improve predictions.

- **Compare Models:**

Use some more complex models. Test Random Forest, XGBoost, and neural networks alongside the current linear model, using cross-validation to pick the one with the best RMSE/MAE/R².

- **Custom LLM:**

given a specific temperature and visitor profile, generates customized recommendations to optimize park visitor comfort under different heat conditions.

Thank You