



# Rideshare: Dynamic Pricing Project: Final Results

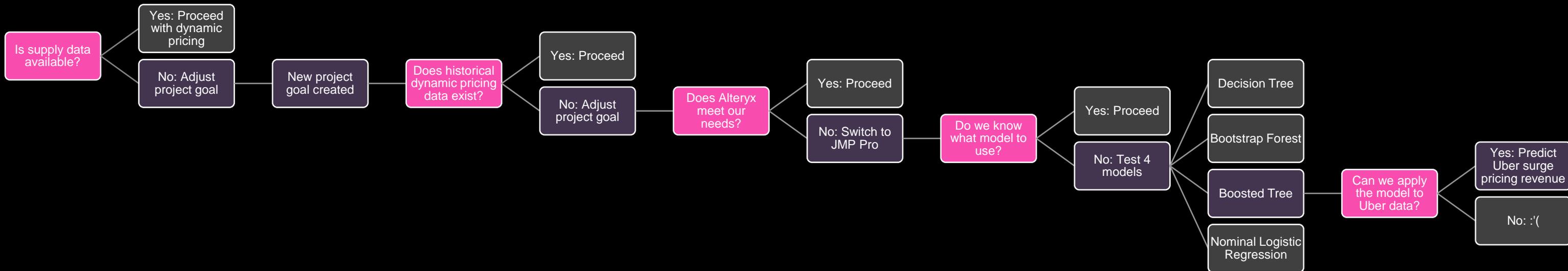
MKT 6V99.501: AI IN MARKETING  
PRESENTED BY: FALGUNI DAS

# Problem Definition

We want to help rideshare companies utilize historical data to predict when surge pricing can be used to maximize revenue and ensure current demand is met with quality, efficient service.



# Our Journey: A Tree of Failures & Learnings



# Data Preparation-Alteryx



1

## Data Cleaning

- Renamed column headers for clarity
- Removed taxi data to keep the focus on rideshare
- Created time of day columns
- Designated base price vs. surge price

2

## Researching Events & Neighborhoods

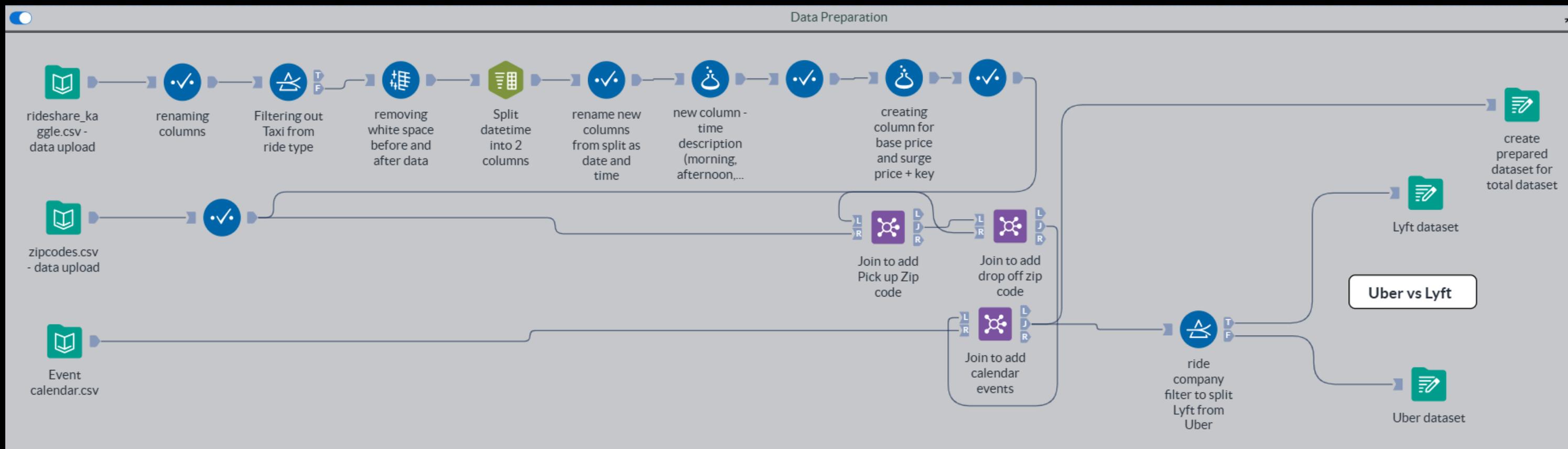
- Established criteria for events
- Researched events and zip codes via government websites & XYZ
- Cleaned data and created keys to join into datasheet

3

## Joining 3 Datasheets

- CSV with original datasheet from Kaggle
- CSV of zip codes and neighborhoods
- CSV of largescale events with keys

# Alteryx Workflow for Data Preparation



# Lyft Model Preparation - JMP Pro

1

## Data Type Identification

- Identified 53 Continuous Variables
- Identified 1 Ordinal Variable
- Identified 14 Nominal Variables

2

## Variable Validation

- Checked for Multicollinearity
  - removed 43 continuous variables
- Removed 7 nominal variables due to data duplication
- Added 1 Boolean Variable
  - Surge (Y=1, N=0)

3

## Data Transformation

- Checked for outliers
- Confirmed normal distribution
  - Transformed skewed variables
    - Log
    - Log + 1

4

## Training Dataset

- Split the data into 3 sets
  - 50% Training Set
  - 30% Validation Set
  - 20% Test Set

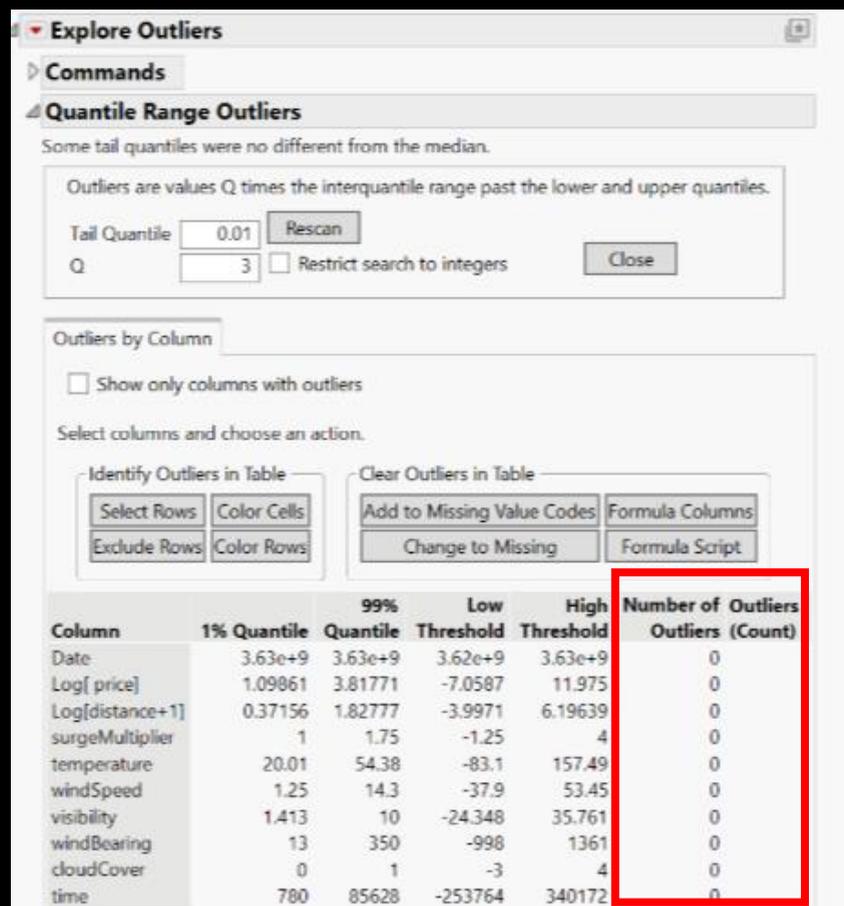
# JMP Pro Model Preparation - Multicollinearity

Multivariate

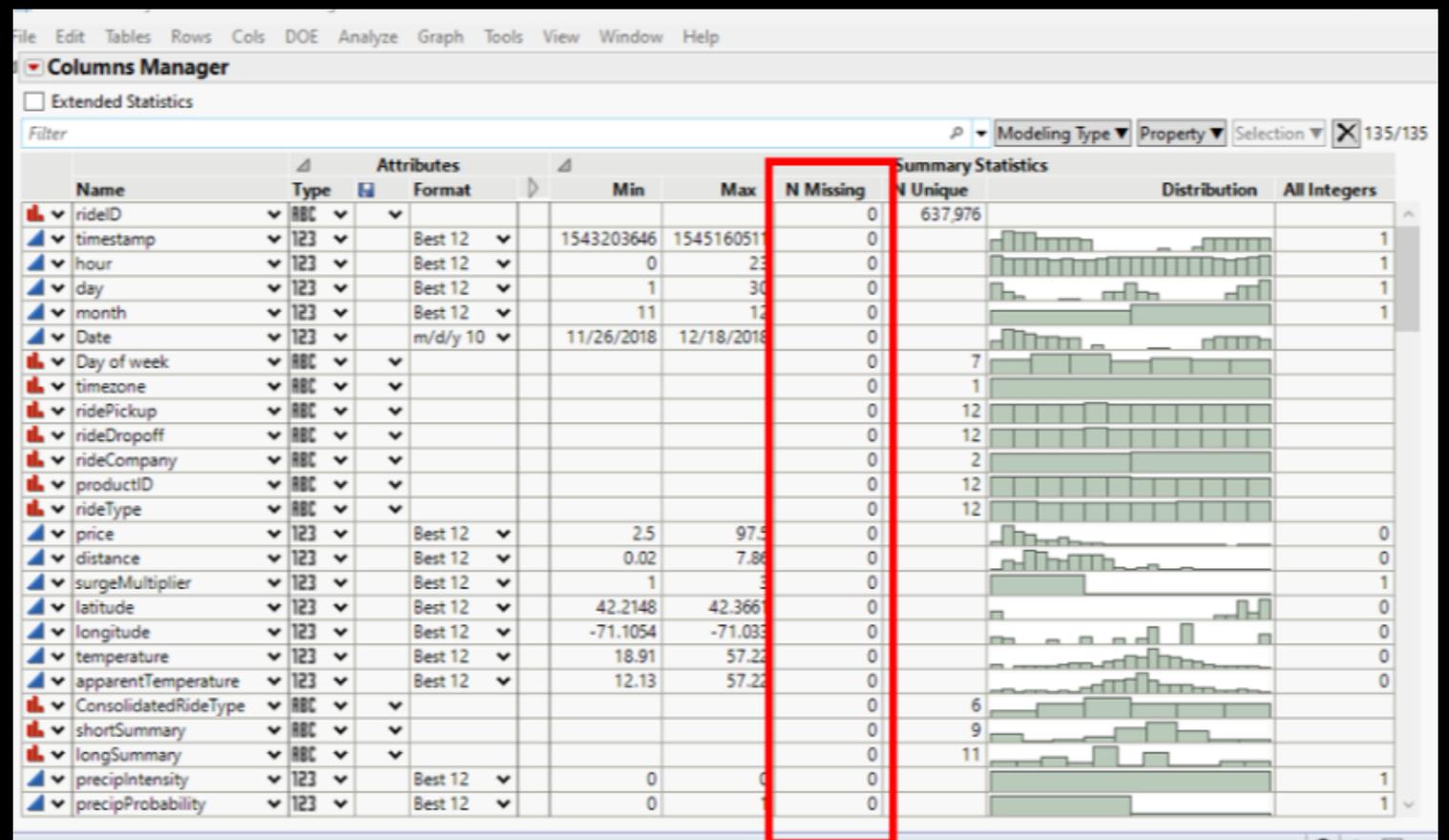
Correlations

	timestamp	hour	day	month	Date	price	distance[Log(distance+1)]	surgeMultiplier	latitude	longitude	temperature	apparentTemperature	precipIntensity	precipProbability	humidity	windSpeed	windGust	windGustTime	visibility	tempen
timestamp	1.0000	-0.0308	-0.3421	0.7704	0.9995	0.0019	0.0041	0.0037	0.0011	0.1688	-0.1349	-0.2478	-0.2346	0.0000	-0.1800	-0.1662	-0.0882	0.0034	0.9988	0.1603
hour	-0.0308	1.0000	0.0683	-0.8625	-0.0670	0.0001	0.0012	-0.0004	0.0196	-0.0023	0.2213	0.1989	0.0000	-0.1584	-0.2743	0.0758	0.0796	-0.0308	0.1716	
day	-0.3421	0.0683	1.0000	-0.8613	-0.3428	-0.0002	-0.0004	-0.0048	0.0026	0.0070	0.0804	-0.0879	-0.2231	0.0000	0.0162	-0.0925	0.4398	0.4025	-0.3352	0.1064
month	0.7704	-0.0625	-0.8613	1.0000	0.7720	0.0025	0.0058	-0.0012	0.0649	-0.1272	-0.0799	0.0271	0.0000	-0.1109	-0.0234	-0.3307	-0.2722	0.7664	0.0112	
Date	0.9995	-0.0670	-0.3420	0.7720	1.0000	0.0019	0.0040	0.0026	0.0011	0.1698	-0.1346	-0.2354	-0.2315	0.0000	-0.1823	-0.1590	-0.0708	0.0005	0.9982	0.1538
price	0.0019	0.0001	-0.0002	0.0025	0.0019	1.0000	0.3616	0.3520	0.3082	0.0025	-0.0019	-0.0020	-0.0016	0.0000	-0.0011	-0.0012	-0.0014	-0.0016	0.0020	0.0022
distance	0.0041	0.0012	-0.0004	0.0058	0.0040	0.3616	1.0000	0.9812	0.0401	-0.0006	-0.0017	-0.0043	-0.0002	0.0000	-0.0010	-0.0035	-0.0016	-0.0022	0.0043	0.0036
Log(distance+1)	0.0037	-0.0019	-0.0049	0.0053	0.0036	0.3520	0.9812	1.0000	0.0400	-0.0012	-0.0004	-0.0046	-0.0002	0.0000	-0.0022	-0.0037	-0.0011	-0.0015	0.0039	0.0042
surgeMultiplier	0.0011	-0.0004	0.0026	-0.0012	0.0011	0.3082	0.0401	1.0000	0.5011	-0.0034	-0.0028	-0.0045	-0.0002	0.0000	-0.0014	-0.0079	0.0046	0.0038	0.0009	0.0021
latitude	0.1688	0.0196	0.0070	0.0649	0.1698	0.0025	-0.0004	-0.0012	1.0000	-0.5384	-0.1081	-0.0981	0.0000	0.0000	-0.0996	-0.1193	-0.0981	0.0229	0.1711	0.1279
longitude	-0.1349	-0.0023	0.0804	-0.1272	-0.1346	-0.0019	-0.0017	-0.0048	-0.0024	-0.5384	1.0000	0.0133	-0.0042	0.0000	0.1379	0.0820	0.0908	-0.0066	-0.1350	-0.1000
temperature	-0.2478	0.2213	-0.0879	-0.0799	-0.2354	-0.0002	-0.0043	-0.0046	-0.0028	-0.1081	0.0133	1.0000	0.9461	0.0000	0.1977	0.3105	0.0596	-0.0088	-0.2471	-0.3248
apparentTemperature	-0.2346	0.1989	-0.2251	0.0271	-0.2315	-0.0015	-0.0032	-0.0036	-0.0045	-0.0981	-0.0042	0.9461	1.0000	0.0000	0.1224	0.3540	-0.2446	-0.2874	-0.2221	-0.3009
precipIntensity	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
precipProbability	-0.1884	-0.1594	0.0182	-0.1109	-0.1823	-0.0011	-0.0020	-0.0022	-0.0014	-0.0036	0.1379	0.1977	0.1224	0.0000	1.0000	0.4379	0.2387	0.1797	-0.1848	-0.6479
humidity	-0.1662	-0.2743	-0.0925	-0.0234	-0.1584	-0.0012	-0.0002	-0.0037	-0.0019	-0.1193	0.0905	0.3105	0.1379	0.4379	1.0000	-0.2099	-0.3085	-0.1801	-0.6485	
windSpeed	0.0758	0.0796	0.4398	-0.3352	-0.0708	-0.0014	-0.0016	-0.0011	0.0046	-0.0981	0.0908	0.0596	-0.2446	0.0000	0.2387	-0.2099	1.0000	0.9381	-0.0738	-0.0338
windGust	0.0034	0.0036	0.4025	-0.2722	0.0005	-0.0016	-0.0019	-0.0015	0.0038	0.0229	0.0066	-0.0088	-0.2446	0.0000	0.1997	-0.3080	0.9381	1.0000	-0.0003	0.0434
windGustTime	0.9988	-0.0309	-0.3352	0.7664	0.9982	0.0020	0.0043	0.0039	0.0009	0.1711	-0.1350	-0.2471	-0.2221	0.0000	-0.1848	-0.1601	-0.0750	-0.0020	1.0000	0.1538
visibility	0.1603	0.1716	0.1064	0.0112	0.1538	0.0022	0.0036	0.0042	0.0021	0.1279	-0.1000	-0.3248	-0.3009	0.0000	-0.6479	-0.6985	-0.0318	0.0434	0.1538	1.0000
temperatureHigh	-0.2262	0.0004	-0.2863	0.0714	-0.2260	-0.0011	-0.0019	-0.0023	-0.0037	-0.1172	0.0066	0.7872	0.8205	0.0000	0.1479	0.4275	-0.1887	-0.2303	-0.2194	-0.2520
Log(temperatureHigh)	-0.2960	0.0080	-0.2430	0.0036	-0.2958	-0.0010	-0.0019	-0.0022	-0.0034	-0.1184	0.0137	0.7946	0.8235	0.0000	0.1396	0.4518	-0.1825	-0.2341	-0.2362	-0.2919
temperatureHighTime	0.9993	-0.0306	-0.3419	0.7713	0.9987	0.0019	0.0041	0.0038	0.0010	0.1722	-0.1345	-0.2593	-0.2316	0.0000	-0.1885	-0.1629	-0.0773	-0.0002	0.9993	0.1537
temperatureLow	-0.3643	0.0157	-0.0880	-0.1387	-0.3642	0.0007	0.0004	0.0008	-0.0049	-0.0390	0.0780	0.5026	0.5725	0.0000	0.2451	0.4719	-0.2364	-0.3408	-0.3482	-0.2668
Log(temperatureLow+1)	-0.3963	0.0236	-0.0617	-0.1726	-0.3964	0.0008	0.0019	0.0012	-0.0046	-0.0318	0.0821	0.5131	0.5880	0.0000	0.2448	0.4610	-0.2394	-0.3242	-0.3196	-0.2382
temperatureLowTime	0.9992	-0.0306	-0.3324	0.7647	0.9985	0.0018	0.0039	0.0038	0.0011	0.1695	-0.1338	-0.2534	-0.2321	0.0000	-0.1888	-0.1660	-0.0622	0.0104	0.9990	0.1544
apparentTemperatureHigh	-0.2135	0.0060	-0.4006	0.1556	-0.2134	-0.0002	-0.0017	-0.0022	-0.0040	-0.0038	-0.0025	0.7094	0.8107	0.0000	0.1307	0.4490	-0.3783	-0.4158	-0.2072	-0.2079
apparentTemperatureLow	0.9993	-0.0308	-0.3421	0.7713	0.9987	0.0019	0.0041	0.0037	0.0010	0.1718	-0.1341	-0.2578	-0.2324	0.0000	-0.1895	-0.1638	-0.0749	-0.0013	0.9992	0.1569
apparentTemperatureHighTime	-0.3989	0.0180	-0.1941	-0.0643	-0.3987	0.0016	0.0040	0.0025	-0.0053	0.0008	0.0173	0.3813	0.4942	0.0000	0.2349	0.3389	-0.2841	-0.4034	-0.3827	-0.2008
apparentTemperatureLowTime	-0.2917	-0.0075	-0.1024	0.7600	0.9985	0.0016	0.0039	0.0038	0.0011	0.1709	-0.1353	-0.2526	-0.2296	0.0000	-0.1875	-0.1629	-0.0670	0.0052	0.9990	0.1532
devPoint	0.5354	-0.0717	-0.4436	0.5904	0.5370	0.0034	0.0076	0.0070	-0.0047	0.1573	-0.0989	-0.3387	-0.3150	0.0000	-0.1400	-0.1321	-0.5748	-0.3145	0.5442	0.2263
pressure	-0.1398	0.0377	0.0917	-0.1386	-0.1380	-0.0006	-0.0016	-0.0005	0.0036	-0.0342	-0.0132	-0.2418	-0.2916	0.0000	-0.4228	-0.3555	0.1155	0.1708	-0.1538	0.4033
windBearing	-0.0833	0.0279	0.0458	-0.0191	-0.0882	-0.0002	-0.0017	-0.0018	-0.0029	-0.0644	0.0762	0.3530	0.2927	0.0000	0.2962	0.4769	0.1593	0.1045	-0.0835	-0.4767
cloudCover	-0.0214	0.3329	-0.0125	-0.0128	-0.0293	-0.0004	0.0027	0.0028	-0.0039	0.0114	0.0083	0.1397	0.1382	0.0000	-0.0833	-0.2058	0.0462	0.0906	-0.0271	0.1171
uvIndex	0.1603	0.1716	0.1064	0.0112	0.1538	0.0022	0.0036	0.0042	0.0021	0.1279	-0.1000	-0.3248	-0.3009	0.0000	-0.6479	-0.6985	-0.0318	0.0434	0.1538	1.0000
visibility1	0.1603	0.1716	0.1064	0.0112	0.1538	0.0022	0.0036	0.0042	0.0021	0.1279	-0.1000	-0.3248	-0.3009	0.0000	-0.6479	-0.6985	-0.0318	0.0434	0.1538	1.0000
ozone	0.1834	0.0345	0.3779	-0.1320	0.1918	-0.0009	-0.0018	-0.0012	0.0052	0.0406	0.0207	-0.2887	-0.4404	0.0000	-0.2361	-0.4236	0.5466	0.5964	0.1878	0.2638
sunriseTime	0.9994	-0.0303	-0.3399	0.7698	0.9987	0.0019	0.0041	0.0037	0.0011	0.1702	-0.1344	-0.2541	-0.2301	0.0000	-0.1878	-0.1624	-0.0705	0.0022	0.9994	0.1543
sunriseTime	0.9994	-0.0303	-0.3399	0.7698	0.9987	0.0019	0.0041	0.0037	0.0011	0.1701	-0.1344	-0.2541	-0.2301	0.0000	-0.1878	-0.1624	-0.0704	0.0023	0.9994	0.1543
moonPhase	-0.8493	0.0035	-0.0887	-0.4506	-0.8457	-0.0020	-0.0047	-0.0041	-0.0006	-0.1985	0.1025	0.3448	0.3252	0.0000	0.1775	0.1477	0.0429	0.0069	-0.8547	-0.1756
precipIntensityMax	-0.2106	0.0008	0.1485	-0.2132	-0.2110	-0.0013	-0.0031	-0.0037	-0.0002	-0.0461	0.0293	0.3316	0.2555	0.0000	0.4063	0.1593	0.2115	0.0989	-0.2007	-0.5440
uvIndexTime	0.9994	-0.0303	-0.3400	0.7699	0.9987	0.0019	0.0041	0.0037	0.0011	0.1703	-0.1345	-0.2536	-0.2298	0.0000	-0.1871	-0.1616	-0.0701	0.0023	0.9994	0.1538
temperatureMin	-0.2963	0.0178	0.0353	-0.1895	-0.2964	-0.0002	-0.0048	-0.0050	-0.0014	-0.1347	0.0445	0.7884	0.8987	0.0000	0.2219	0.4590	0.1740	0.0822	-0.2820	-0.3334
temperatureMinTime	0.9982	-0.0307	-0.3329	0.7645	0.9975	0.0018	0.0039	0.0035	0.0011	0.1679	-0.1383	-0.2478	-0.2281	0.0000	-0.1899	-0.1691	-0.0535	0.0211	0.9983	0.1540
temperatureMax	-0.1845	0.0044	-0.3074	0.1083	-0.1844	-0.0015	-0.0019	-0.0023	-0.0037	-0.1111	-0.0079	0.7791	0.7915	0.0000	0.1440	0.3923	-0.1207	-0.1440	-0.1780	-0.2984
temperatureMaxTime	0.9993	-0.0308	-0.3433	0.7722	0.9986	0.0019	0.0041	0.0037	0.0010	0.1727	-0.1350	-0.2530	-0.2262	0.0000	-0.1889	-0.1595	-0.0603	-0.0082	0.9993	0.1581
apparentTemperatureMin	-0.3771	0.0171	-0.0250	-0.1883	-0.3771	-0.0015	-0.0042	-0.0042	-0.0026	-0.0965	0.0081	0.8187	0.8186	0.0000	0.2277	0.5102	-0.0198	-0.1126	-0.2703	-0.3433
apparentTemperatureMinTime	0.9983	-0.0305	-0.3339	0.7652	0.9976	0.0018	0.0039	0.0035	0.											

# JMP Pro Model Preparation - Outliers & Missing Data

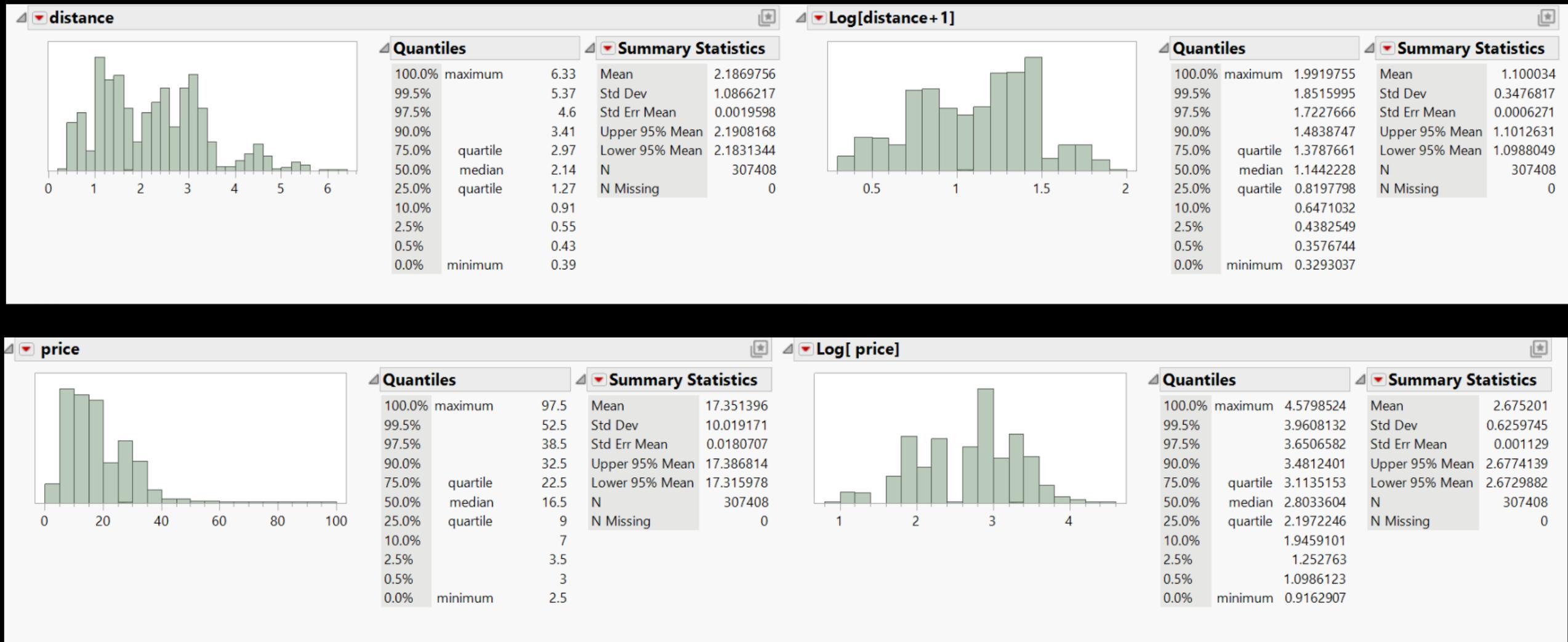


- Of the remaining variables there are no outliers



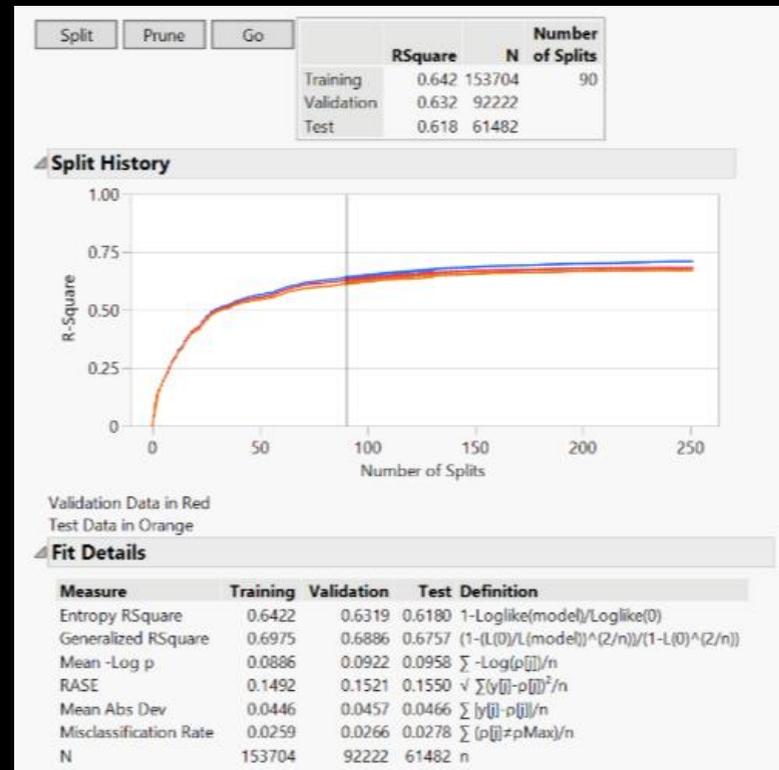
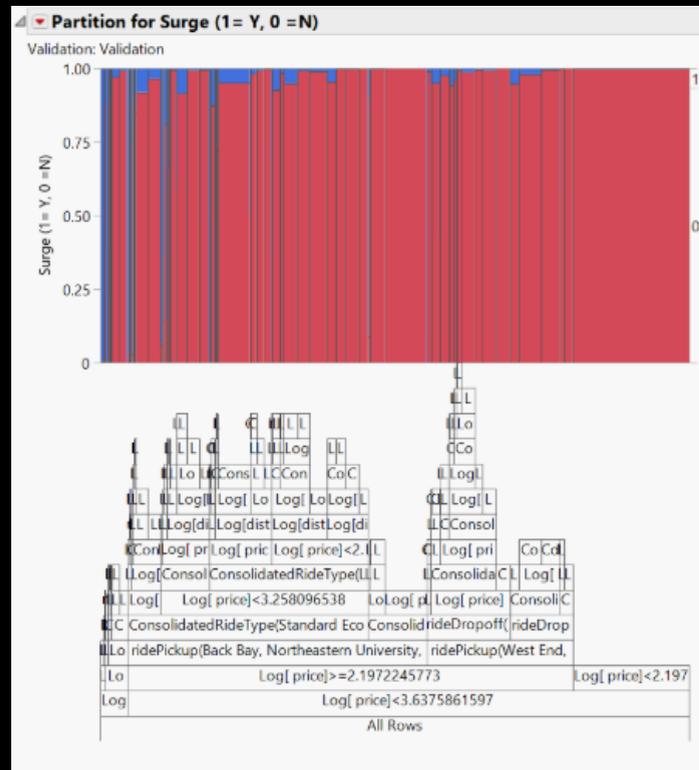
- Additionally, there are no missing data points in any of the variables

# JMP Pro Model Preparation - Transformations



- Transformed Distance and Price because they were showing right skewness
  - Distance – Log +1 transformation
  - Price – Log transformation

# Model 1 - Decision Tree



## Confusion Matrix

Training			Validation			Test		
Actual	Predicted		Actual	Predicted		Actual	Predicted	
Surge (1= Y, 0 =N)	0	1	Surge (1= Y, 0 =N)	0	1	Surge (1= Y, 0 =N)	0	1
0	142546	757	0	85424	458	0	56921	327
1	3231	7170	1	1996	4344	1	1382	2852

Training			Validation			Test		
Actual	Predicted		Actual	Predicted		Actual	Predicted	
Surge (1= Y, 0 =N)	0	1	Surge (1= Y, 0 =N)	0	1	Surge (1= Y, 0 =N)	0	1
0	0.995	0.005	0	0.995	0.005	0	0.994	0.006
1	0.311	0.689	1	0.315	0.685	1	0.326	0.674

## Model Results (Test data):

- Model Type = Decision Tree – 90 splits or nodes
- Entropy R-Squared = 0.6180
- Misclassification rate = 0.0278
- Confusion Matrix = correctly predicted 56921 “No” values + 2852 “Yes” values = 59773 correct predictions = 0.9722



# Model 2 - Bootstrap Forest

**Bootstrap Forest for Surge (1 = Y, 0 = N)**

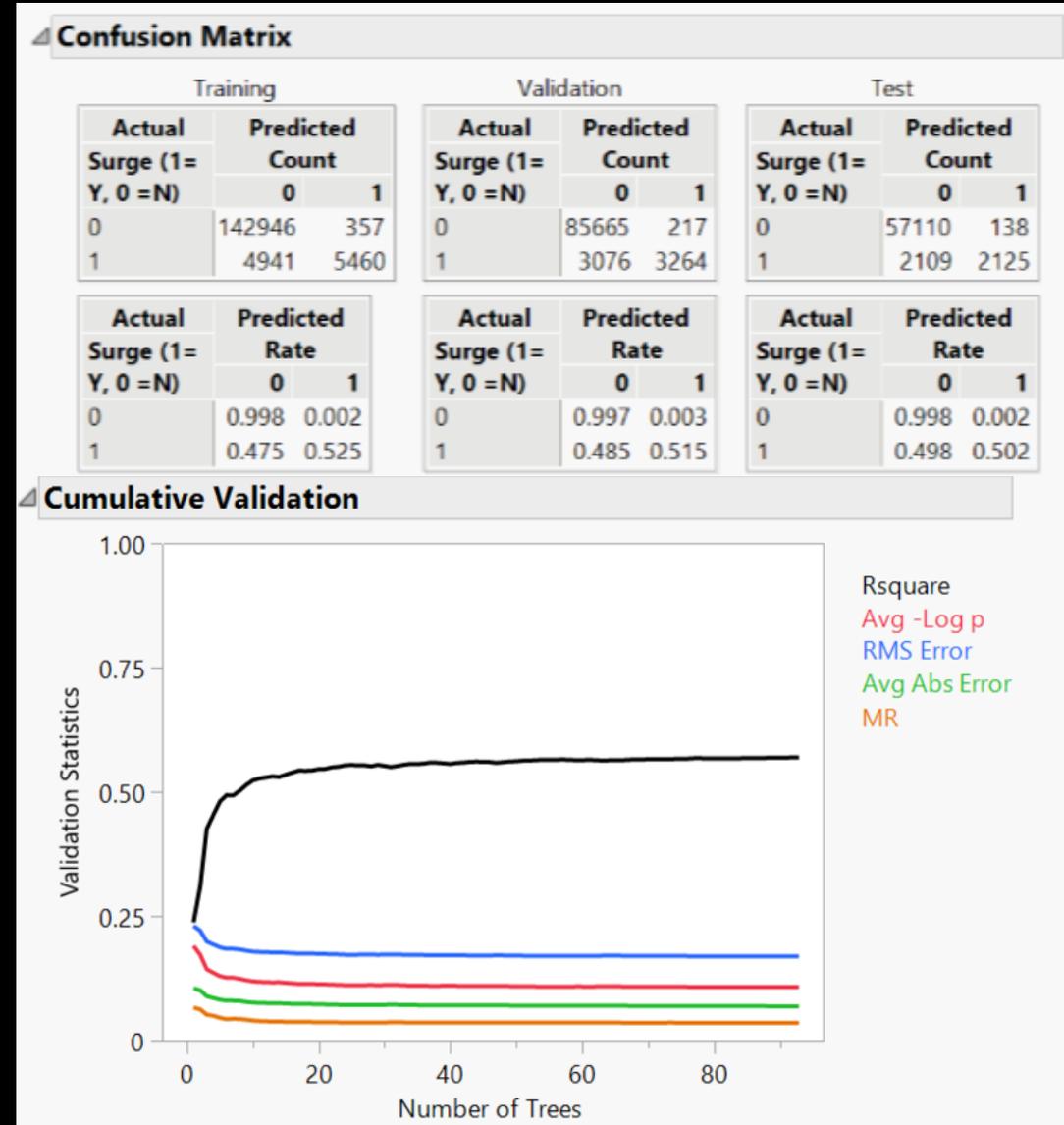
**Specifications**

Target	Surge (1 = Y, 0 = N)	Training Rows:	153704
Validation Column:	Validation	Validation Rows:	92222
		Test Rows:	61482
Number of Trees in the Forest:	500	Number of Terms:	18
Number of Terms Sampled per Split:	11	Bootstrap Samples:	153704
		Minimum Splits per Tree:	10
		Minimum Size Split:	307

**Overall Statistics**

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.5827	0.5701	0.5634	$1 - \text{Loglike}(\text{model}) / \text{Loglike}(0)$
Generalized RSquare	0.6418	0.6305	0.6241	$(1 - L(0) / L(\text{model}))^{(2/n)} / (1 - L(0))^{(2/n)}$
Mean -Log p	0.1033	0.1076	0.1094	$\sum -\text{Log}(\rho_{ij}) / n$
RASE	0.1665	0.1697	0.1718	$\sqrt{\sum (y_{ij} - \rho_{ij})^2 / n}$
Mean Abs Dev	0.0678	0.0692	0.0696	$\sum  y_{ij} - \rho_{ij}  / n$
Misclassification Rate	0.0345	0.0357	0.0365	$\sum (\rho_{ij} \neq \rho_{\text{Max}}) / n$
N	153704	92222	61482	n

**Number of Trees**  
93



## Model Results (Test data):

- Model Type = Bootstrap Forest – 93 trees
- Entropy R-Squared = 0.5634
- Misclassification rate = 0.0365
- Confusion Matrix = correctly predicted 57110 “No” values + 2125 “Yes” values = 59235 correct predictions = 0.9635

# Model 3 - Boosted Tree

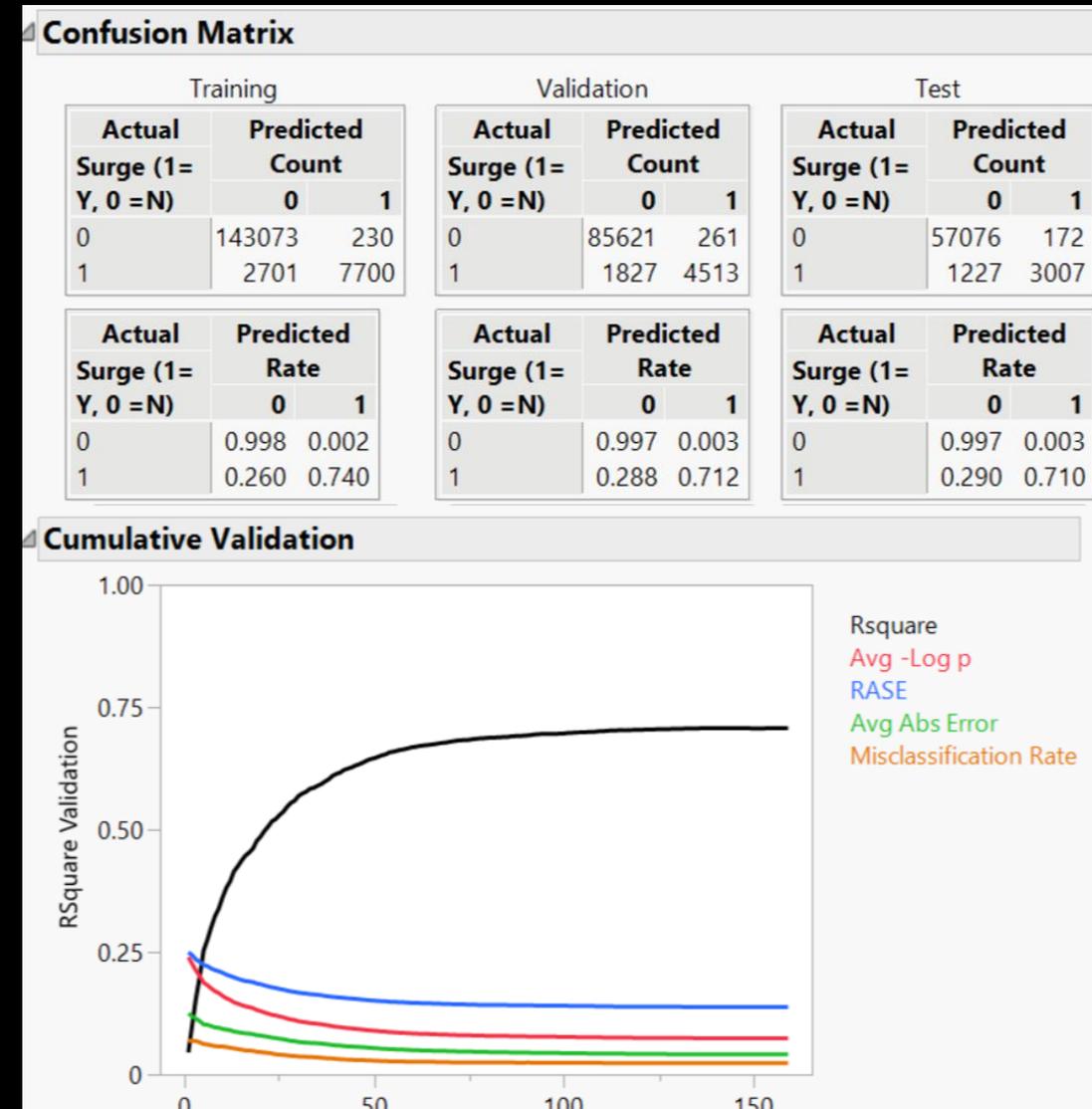
**Boosted Tree for Surge (1 = Y, 0 = N)**

**Specifications**

Target: Surge (1 = Y, 0 = N)    Number of training rows: 153704  
 Validation Column: Validation    Number of validation rows: 92222  
 Number of Layers: 159    Number of test rows: 61482  
 Splits per Tree: 20  
 Learning Rate: 0.199  
 Overfit Penalty: 0.0001

**Overall Statistics**

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.7412	0.7063	0.7008	$1 - \text{Loglike}(\text{model}) / \text{Loglike}(0)$
Generalized RSquare	0.7866	0.7562	0.7514	$(1 - L(0) / L(\text{model}))^{2/n} / (1 - L(0)^{2/n})$
Mean -Log p	0.0641	0.0735	0.0750	$\sum -\text{Log}(\rho[j]) / n$
RASE	0.1272	0.1374	0.1389	$\sqrt{\sum (y[j] - \rho[j])^2 / n}$
Mean Abs Dev	0.0373	0.0405	0.0409	$\sum  y[j] - \rho[j]  / n$
Misclassification Rate	0.0191	0.0226	0.0228	$\sum (\rho[j] \neq \rho_{\text{Max}}) / n$
N	153704	92222	61482	n



## Model Results (Test data):

- Model Type = Boosted Tree – 159 layers
- Entropy R-Squared = 0.7008
- Misclassification rate = 0.0228
- Confusion Matrix = correctly predicted 57076 “No” values + 3007 “Yes” values = 60083 correct predictions = 0.9772

# Model 4 - Regression - Nominal Logistic

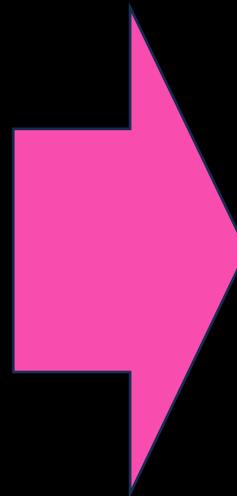
Nominal Logistic Fit for Surge (1= Y, 0 =N)

Effect Summary

Source	Logworth	PValue
Log[ price]	2590.960	0.00000
ConsolidatedRideType	2457.180	0.00000
Log[distance+1]	1888.654	0.00000
ridePickup	525.597	0.00000
rideDropoff	261.969	0.00000
cloudCover	3.331	0.00047
shortSummary	2.833	0.00147
windBearing	1.927	0.01182
Day of week	1.382	0.04147
uvIndex	1.018	0.09592
Date	0.804	0.15714
timeDescription	0.716	0.19246
windSpeed	0.429	0.37258
visibility	0.226	0.59461
EventLocation_Arrival	0.194	0.63945
time	0.186	0.65228
temperature	0.131	0.73883
EventDescription_Arrival	0.000	0.99956

Fit Details

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.5898	0.5853	0.5720	$1 - \text{Loglike}(\text{model}) / \text{Loglike}(0)$
Generalized RSquare	0.6486	0.6449	0.6323	$(1 - (L(0)/L(\text{model}))^{2/n}) / (1 - L(0)^{2/n})$
Mean -Log p	0.1015	0.1038	0.1073	$\sum -\text{Log}(\rho[j]) / n$
RASE	0.1564	0.1579	0.1610	$\sqrt{\sum (y[j] - \rho[j])^2 / n}$
Mean Abs Dev	0.0524	0.0529	0.0539	$\sum  y[j] - \rho[j]  / n$
Misclassification Rate	0.0283	0.0287	0.0299	$\sum (\rho[j] \neq \rho_{\text{Max}}) / n$
N	153704	92222	61482	n



Nominal Logistic Fit for Surge (1= Y, 0 =N)

Effect Summary

Source	Logworth	PValue
Log[ price]	2593.833	0.00000
ConsolidatedRideType	2460.084	0.00000
Log[distance+1]	1890.921	0.00000
ridePickup	526.357	0.00000
rideDropoff	268.013	0.00000
cloudCover	3.855	0.00014
shortSummary	2.936	0.00116
Day of week	1.738	0.01826
windBearing	1.730	0.01861

Fit Details

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.5896	0.5852	0.5725	$1 - \text{Loglike}(\text{model}) / \text{Loglike}(0)$
Generalized RSquare	0.6483	0.6448	0.6328	$(1 - (L(0)/L(\text{model}))^{2/n}) / (1 - L(0)^{2/n})$
Mean -Log p	0.1016	0.1039	0.1072	$\sum -\text{Log}(\rho[j]) / n$
RASE	0.1564	0.1579	0.1608	$\sqrt{\sum (y[j] - \rho[j])^2 / n}$
Mean Abs Dev	0.0524	0.0529	0.0539	$\sum  y[j] - \rho[j]  / n$
Misclassification Rate	0.0284	0.0286	0.0299	$\sum (\rho[j] \neq \rho_{\text{Max}}) / n$
N	153704	92222	61482	n

## Model Results (Test data):

- Model Type = Nominal Logistic – Regression
- Entropy R-Squared = 0.5725
- Misclassification rate = 0.0299
- Classification rate = 0.9701

# Battle of the Models - Model Selection

## Model Summary & Selection

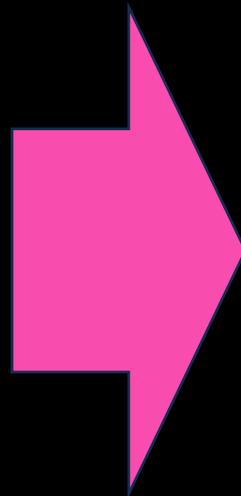
Model = Surge (1=Y, 0=N)	Entropy R-Squared	Missclassification Rate	Classification Rate	# of Nodes/ Trees/Layers
Nominal Logistic	0.5725	3.0%	97.0%	N/A
Decision Tree	0.6180	2.8%	97.2%	90
Bootstrap Forest	0.5634	3.7%	96.3%	93
<b>Boosted Tree</b>	<b>0.7008</b>	<b>2.3%</b>	<b>97.7%</b>	<b>159</b>

### Selection Results (Test data):

- Model = Boosted Tree
  - Highest Entropy R-Squared
  - Lowest Misclassification rate
  - Highest Classification rate

# Model Application - From Lyft to Uber

Surge Rate Distribution - Lyft		
Surge Rate	Record Count by Rate	% Distribution
1.00	663	4%
1.25	6150	39%
1.50	4303	27%
1.75	2362	15%
2.00	2239	14%
2.50	154	1%
3.00	12	0%
<b>Total</b>	<b>15883</b>	<b>100%</b>



Surge Rate Distribution - Uber		
Surge Rate	Record Count by Rate	% Distribution
1.00	478	4%
1.25	4435	39%
1.50	3103	27%
1.75	1703	15%
2.00	1615	14%
2.50	111	1%
3.00	9	0%
<b>Total</b>	<b>11454</b>	<b>100%</b>

## Assumptions:

- Leveraging the Lyft model predicted records with surge pricing
  - Calculated the distribution of records by surge rate
  - Applied the distribution to the model predicted records for Uber

# Model Application - Uber Surge Prediction

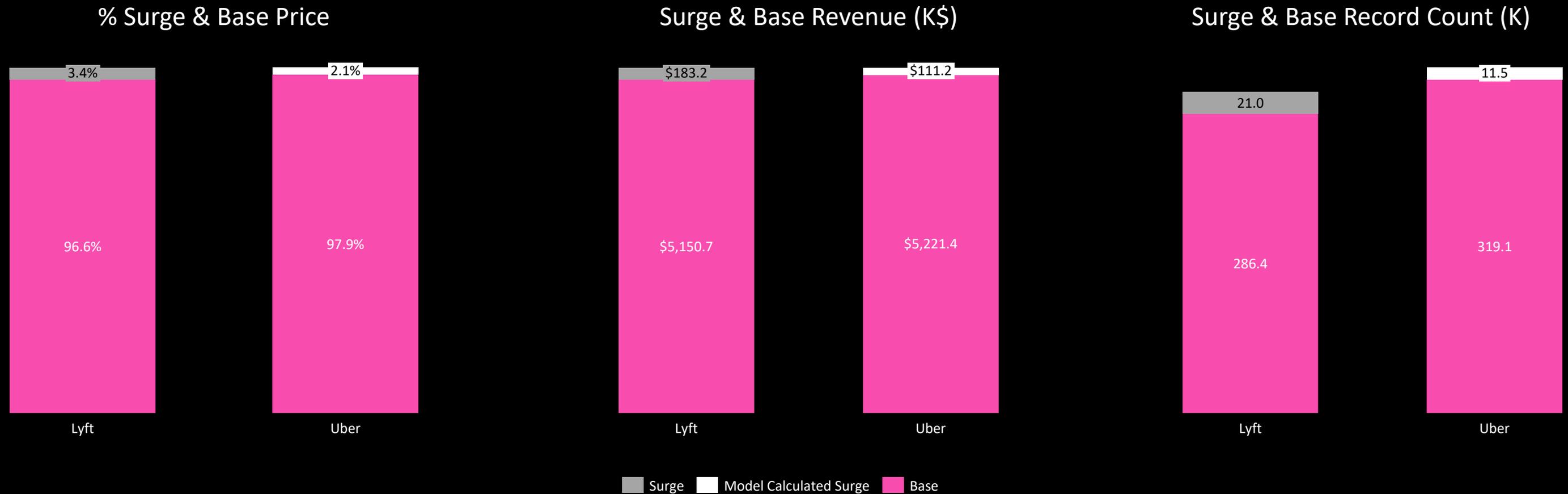
Adding surge pricing can increase Uber's revenue by \$111k or ~2.1%

Surge Rate Distribution & Revenue Estimation - Uber					
Surge Rate	Record Count by Rate	% Distribution	Without Surge	With Surge	Additional Revenue
1.00	478	4%	\$9,269	\$9,269	\$0
1.25	4435	39%	\$85,977	\$107,471	\$21,494
1.50	3103	27%	\$60,156	\$90,234	\$30,078
1.75	1703	15%	\$33,021	\$57,786	\$24,766
2.00	1615	14%	\$31,301	\$62,602	\$31,301
2.50	111	1%	\$2,153	\$5,382	\$3,229
3.00	9	0%	\$168	\$503	\$336
<b>Total</b>	<b>11454</b>	<b>100%</b>	<b>\$222,045</b>	<b>\$333,248</b>	<b>\$111,204</b>

## Assumptions:

- Surge rate distribution applied to the 11454 Uber records that the model predicted would have surge
- Average price for Uber records that are predicted to have surge pricing is \$19.39 or \$222k in revenue
- With the surge rate applied to the average price, the revenue has the potential to increase to \$333k
- Possible additional revenue Uber can recognize is \$111k, a 2.1% increase vs their revenue with no surge

# Model Results - %, Revenue, & Record Count of Surge & Base Price by Ride Share Company



## Comparisons:

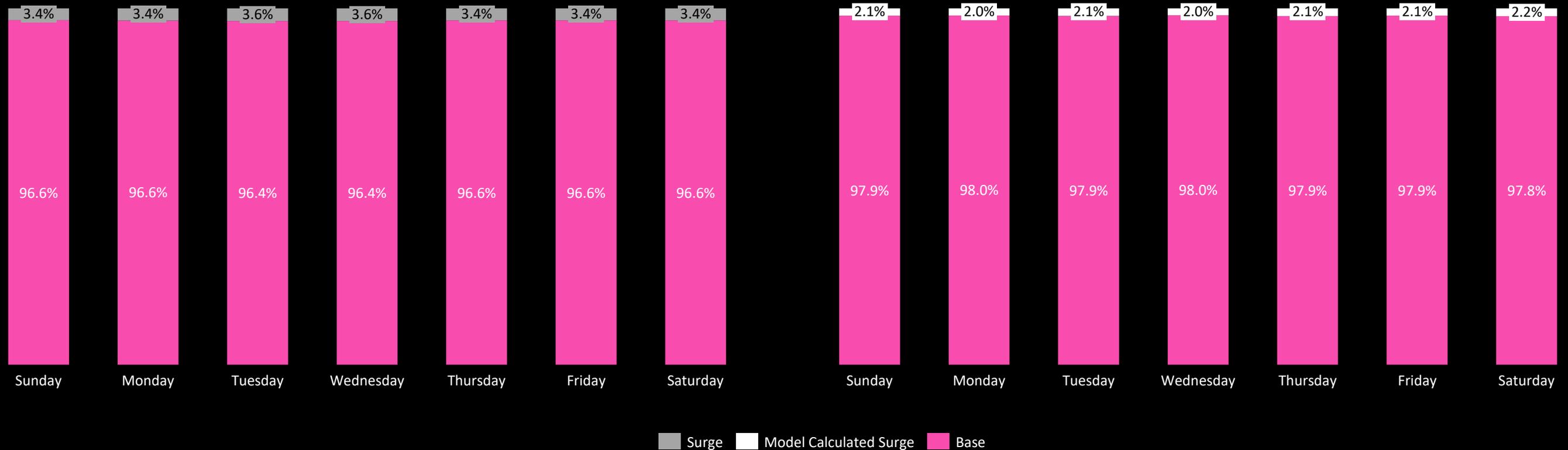
- % Surge and Base Price
- Surge & Base Revenue (K\$)
- Surge & Base Record count (K)

The model surge results for Uber are directionally similar to the actual surge results for Lyft

# Model Results - % Surge & Base Price by Day and Ride Share Company

% Surge & Base Price - Lyft

% Surge & Base Price - Uber



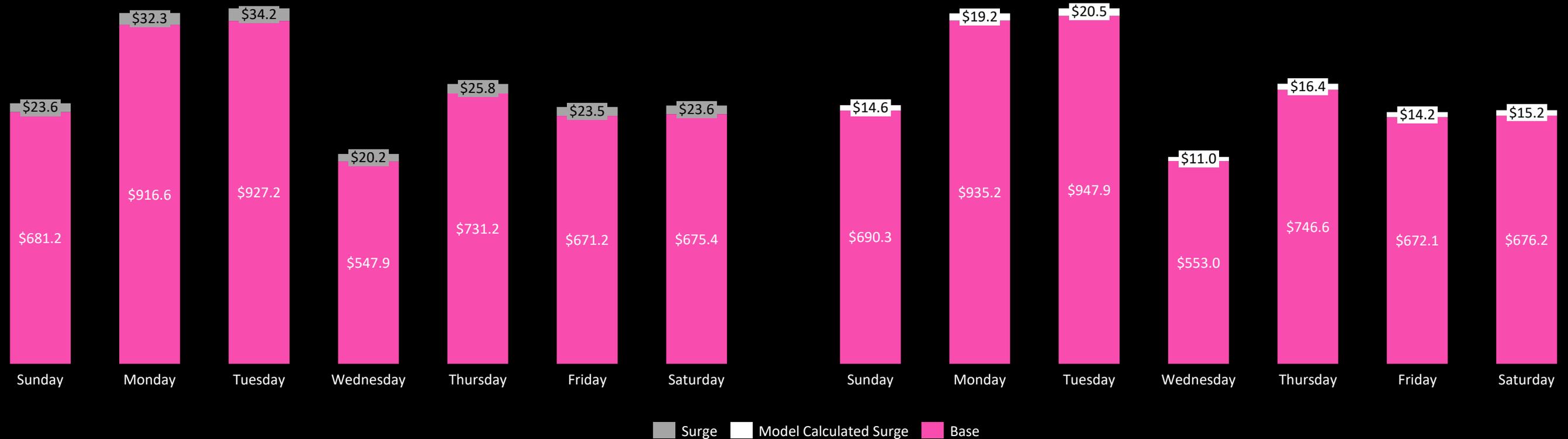
## % Surge & Base Price by Day of the Week

- Looking at the % of surge vs the base price there are no visual signs that applying the model to Uber data creates unreasonable results

# Model Results – Total Revenue – Surge & Base by Day and Ride Share Company

Surge & Base Revenue – Lyft (K\$)

Surge & Base Revenue – Uber (K\$)



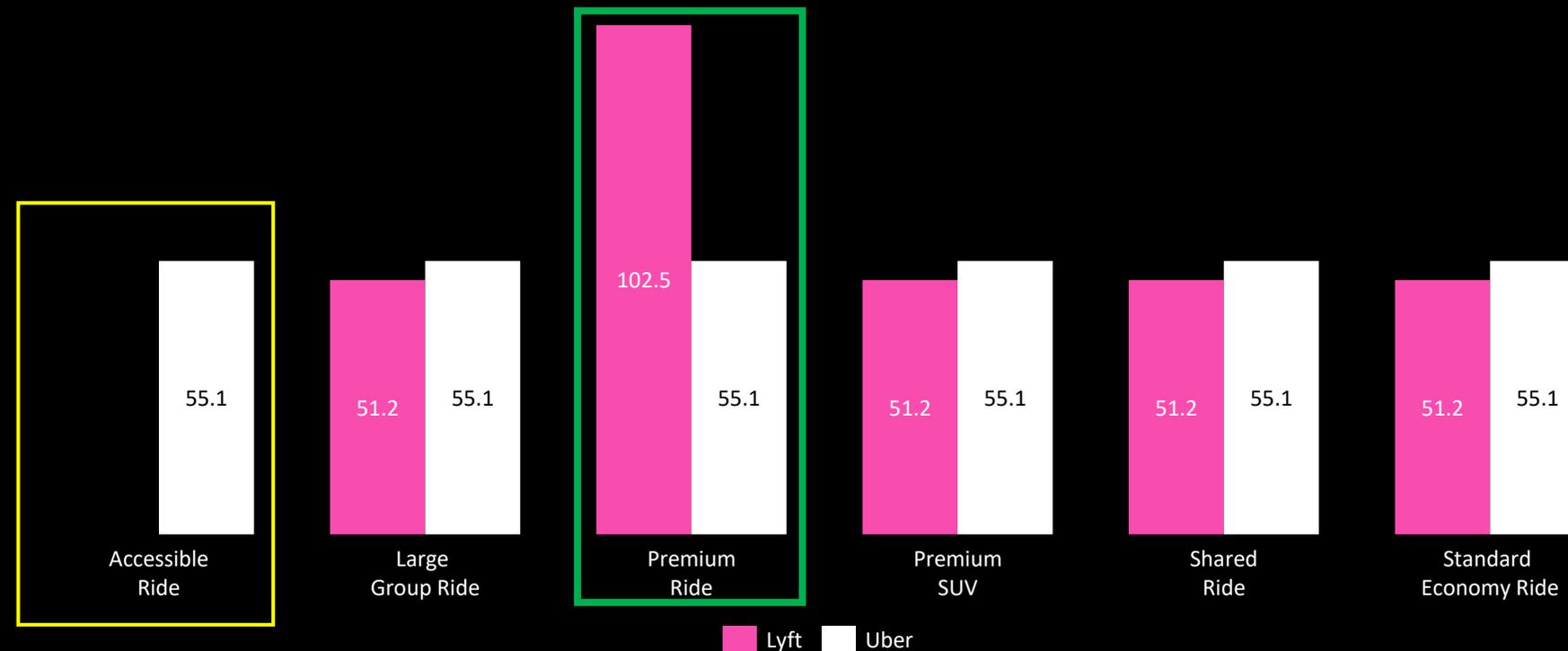
## Surge & Base Revenue by Day of the Week

- Looking at the revenue by surge and base prices there are no visual signs that applying the model to Uber data creates unreasonable results
- However, even if the results do not look unreasonable, they do look like they could be low

What is the driver for the lower model results?

# Model Result Validation – Total Count – Ride by Ride Type & Company

Count of Rides by Ride Type & Company (K)



## Lyft & Uber Ride count by Ride Type (K)

- The primary driver for Lyft's higher level of surge pricing is driven by the higher number of premium rides
- Uber needs to review accessible rides as there is no data in the model because the Lyft data did not have any

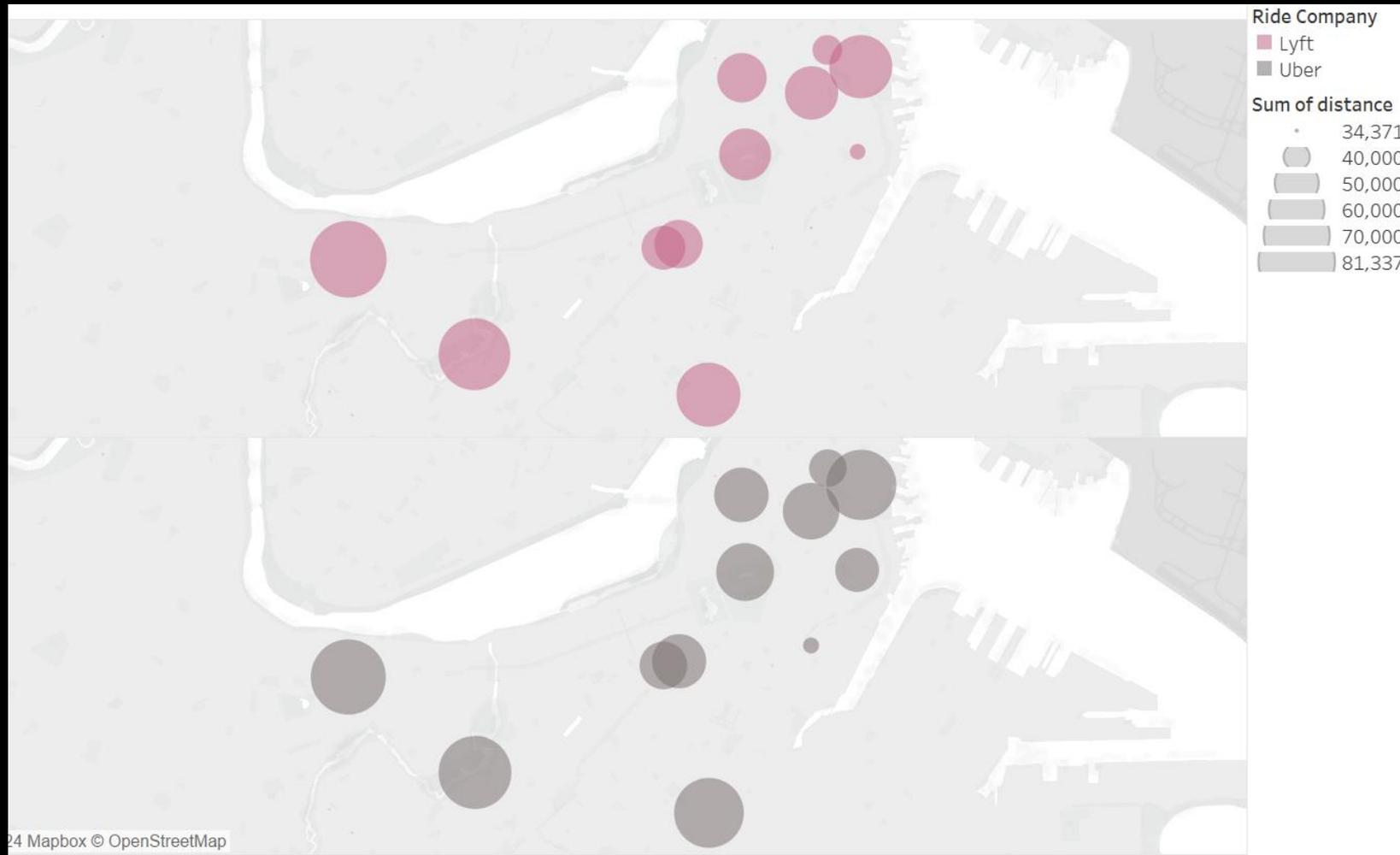
Overall, the model's predictions for Surge pricing for Uber produces results that are in line with expectations based on the data.



# Uber ☆ vs. Lyft ♡

Lyft

Uber



## Total Distance from Pick-up Point

- Assumptions:
  - Dots on the map represent the total distance the driver traveled from the pick-up point
  - The dot location is based on the neighborhood where the pick-up location was.

Though there are minor differences in the two maps, they are visually similar.

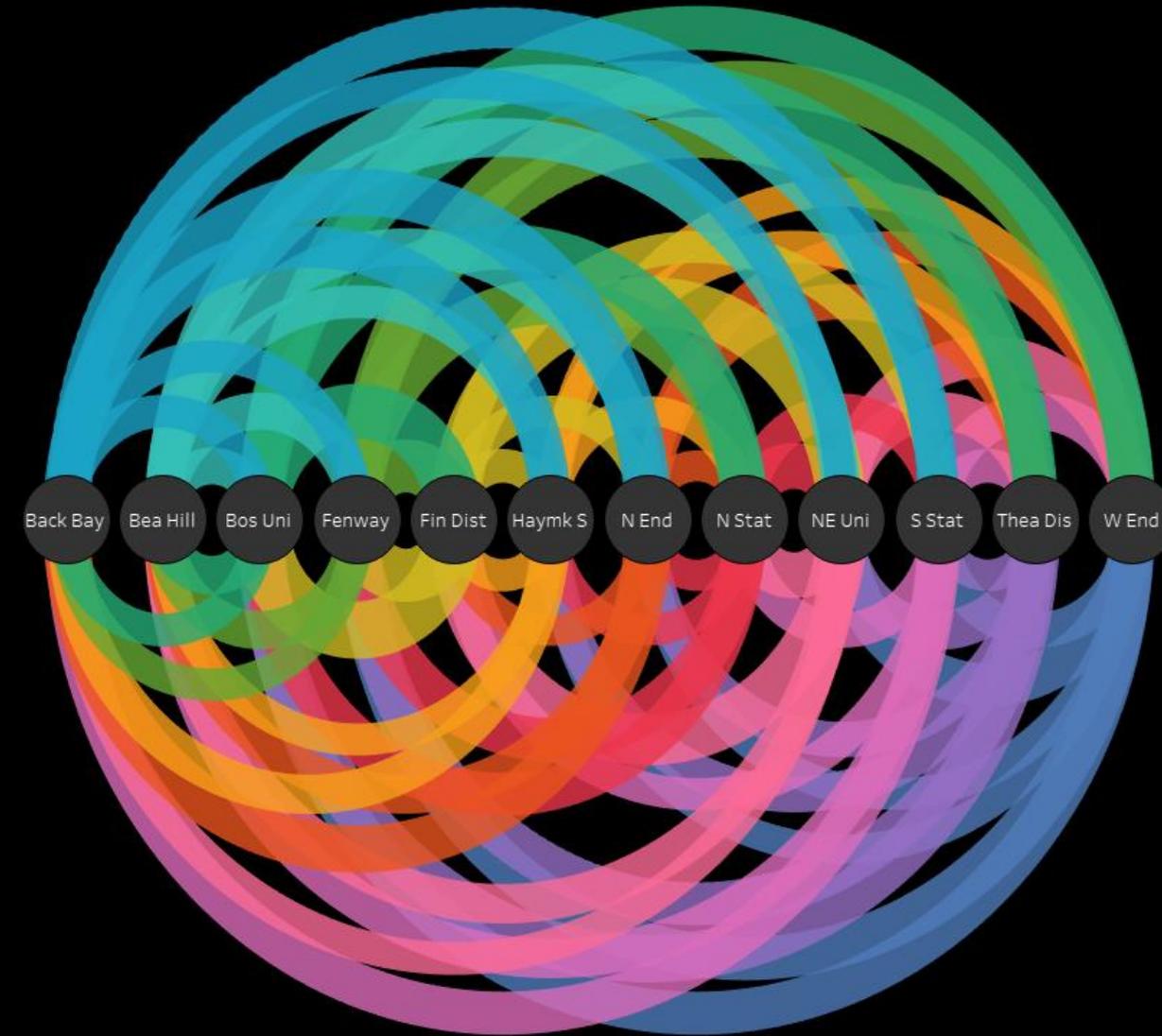
# Finding 5 - Flow Between Pick-up & Drop-off Neighborhoods

## ARC SANKEY

Ride Pick-up Neighborhood
Back Bay
Beacon Hill
Boston Uni
Fenway
Financial Dist
Haymarket Sq
North End
North Stat
NE Uni
South Stat
Theatre Dist
West End



Ride Drop-off Neighborhood					
Boston Uni	Fenway	Haymarket Sq	North End	NE Uni	South Stat
Boston Uni	Fenway	Haymarket Sq	North End	NE Uni	South Stat
Back Bay	Beacon Hill	Financial Dist	North Stat	Theatre Dist	West End
Back Bay	Beacon Hill	Financial Dist	North Stat	Theatre Dist	West End
Boston Uni	Fenway	Haymarket Sq	North End	NE Uni	South Stat
Back Bay	Beacon Hill	Financial Dist	North Stat	Theatre Dist	West End
Back Bay	Beacon Hill	Financial Dist	North Stat	Theatre Dist	West End
Boston Uni	Fenway	Haymarket Sq	North End	NE Uni	South Stat
Back Bay	Beacon Hill	Financial Dist	North Stat	Theatre Dist	West End
Back Bay	Beacon Hill	Financial Dist	North Stat	Theatre Dist	West End
Boston Uni	Fenway	Haymarket Sq	North End	NE Uni	South Stat
Boston Uni	Fenway	Haymarket Sq	North End	NE Uni	South Stat



The data exploration on the connections between ride pick-up neighborhoods and drop-off neighborhoods indicate that every pick-up neighborhood had 6 drop-off neighborhoods. The relationships between pick-up and drop-off neighborhoods can be seen in both the summary table and the Arc Sankey Diagram

Overall, there are small differences in the thicknesses of the arcs between the different ride paths

# Ethical Considerations & Mitigation

## 1 Transparency and Customer Trust

Transparency is vital to maintaining customer trust when implementing surge pricing. Customers should understand the reasons for price increases, such as demand spikes due to weather or local events.

Provide customers separate pricing for base and surge, allows for clear identification of pricing factors.

## 2 Preventing Exploitative Pricing Practices

Unregulated surge pricing during emergencies or high-demand events can lead to customer dissatisfaction and harm the brand's reputation.

Implementing a surge rate constraint to ensure surge pricing remains within reasonable thresholds.

## 3 AI Accountability and Bias Mitigation

Bias in AI-driven pricing can unfairly target specific neighborhoods or times, leading to inequitable outcomes.

We reduced multicollinearity by removing highly correlated variables from the dataset during the model preparation.

## 4 Ethical Use of Public Data

Using weather and event data responsibly ensures the model accurately reflects real-world demand without creating artificial surges.

We cross-referenced event data from multiple sources and filtered it using strict criteria to include only high-impact events likely to influence demand.



# Recommendations for Rideshare Companies



## Implement Surge Pricing

Uber should use our surge pricing model to implement surge pricing, increase revenues and compete with Lyft. Update model for Accessible Ride considerations

## Predict Surge Rates

A second model should be created to use surge predicted records to then predict the surge rate by leveraging a regression model and the dummy variables that were discarded



## Explore Dynamic Pricing

Both rideshare companies have an opportunity to take surge pricing a step further by exploring dynamic pricing models.



# References

1. Mengjing Chen, Weiran Shen, Pingzhong Tang, Song Zuo. "Dispatching Through Pricing: Modeling Ride-Sharing and Designing Dynamic Prices." *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*. International Joint Conferences on Artificial Intelligence, 2019. 165-171.
2. Mobius. *Dynamic Pricing Dataset: Maximize Revenue and Profitability by Dynamic Pricing*. 28 January 2024. Kaggle. September 2024. <[www.kaggle.com/datasets/arashnic/dynamic-pricing-dataset/data](https://www.kaggle.com/datasets/arashnic/dynamic-pricing-dataset/data)>.
3. Mohammad Asghari, Cyrus Shahabi. "ADAPT-Pricing: A Dynamic And Predictive Technique for Pricing to Maximize Revenue in Ridesharing Platforms." *Association for Computing Machinery*. Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, , 2018. 189-198.
4. Munde, Ravi. *Uber Lyft Cab Prices*. 14 November 2019. 14 September 2024. <<https://www.kaggle.com/datasets/ravi72munde/uber-lyft-cab-prices>>.
5. Sahota, Neil. "Harnessing AI For Dynamic Pricing For Your Business." 24 June 2024. *Forbes*.