Analysis on Trips and Driver Performance

Tourmaline Labs Industry Partner Project

Project Lead: Renaldy Herlim

Yishan Li, Samuel Huang, Lawrence Yi

Tourmaline Labs

Data Science Student Society

UC San Diego

HALICIOĞLU DATA SCIENCE INSTITUTE

Agenda Overview

Project Goal
- What are the questions we are trying to solve?
- Who will benefit from this project?

Data Overview and Cleaning
- Source of data
- What does the data look like?

Exploratory Data Analysis
- Surface-level insights
- Transforming the data to see patterns

Analysis of Benchmarks
- Break down the variables
- Track and benchmark the performance of drivers

In-Depth Analysis
- Analyzing the trips data more in depth
- Statistical analyses on the key variables of the datasets

Project Goals & Questions

Main Question: What Key Performance Indicators(KPI) of a driver should an enterprise pay attention to in order to achieve **better safety and reduce costs**?

- What are the **variables we should look out for** in the data?
- How can we create a better **benchmarking** system (comparing a driver to a standard)?
- How do we define this **standard**?
- Which drivers are **exceeding the standards** and expectations?
- Which drivers are **below the minimum standards**?

Datasets and Cleaning

Trips Data

Sample of 120,000 trips data

Data is sampled from San Diego, Denver, and Seattle (40,000 rows for each city)

Variables:
- Trip ID and Driver's ID
- Location and Time of Trip
- Duration and Distance of Trip
- Scores Associated for Trip

Merged

Telematics/Events Data

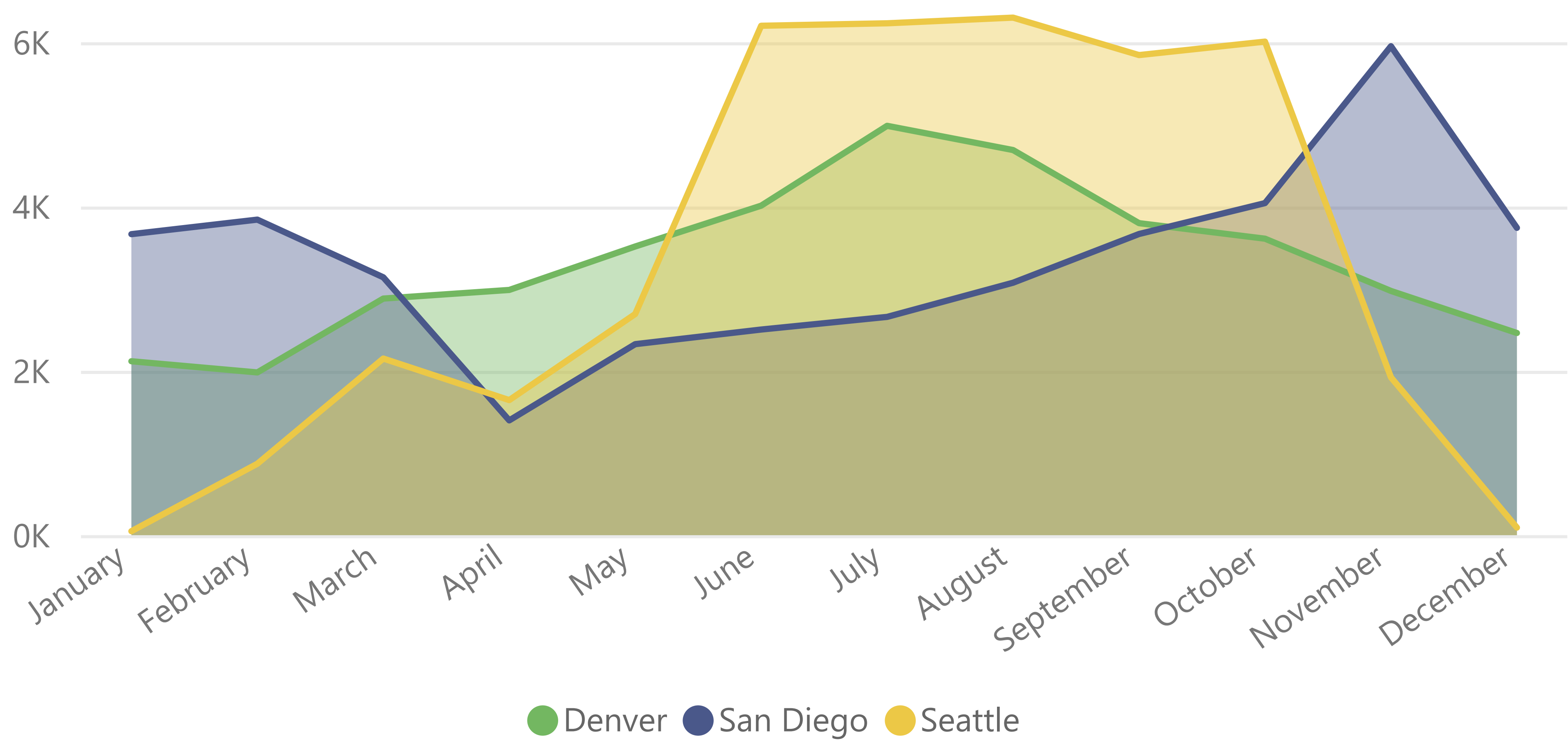
The recorded events associated with the trips

Each trip may be associated with zero or more recorded events

Variables:
- Type of Event (Speeding, Brake, Phone, etc.)
- Condition of Event (City vs Highway)
- Severity score of event
- Location and Time

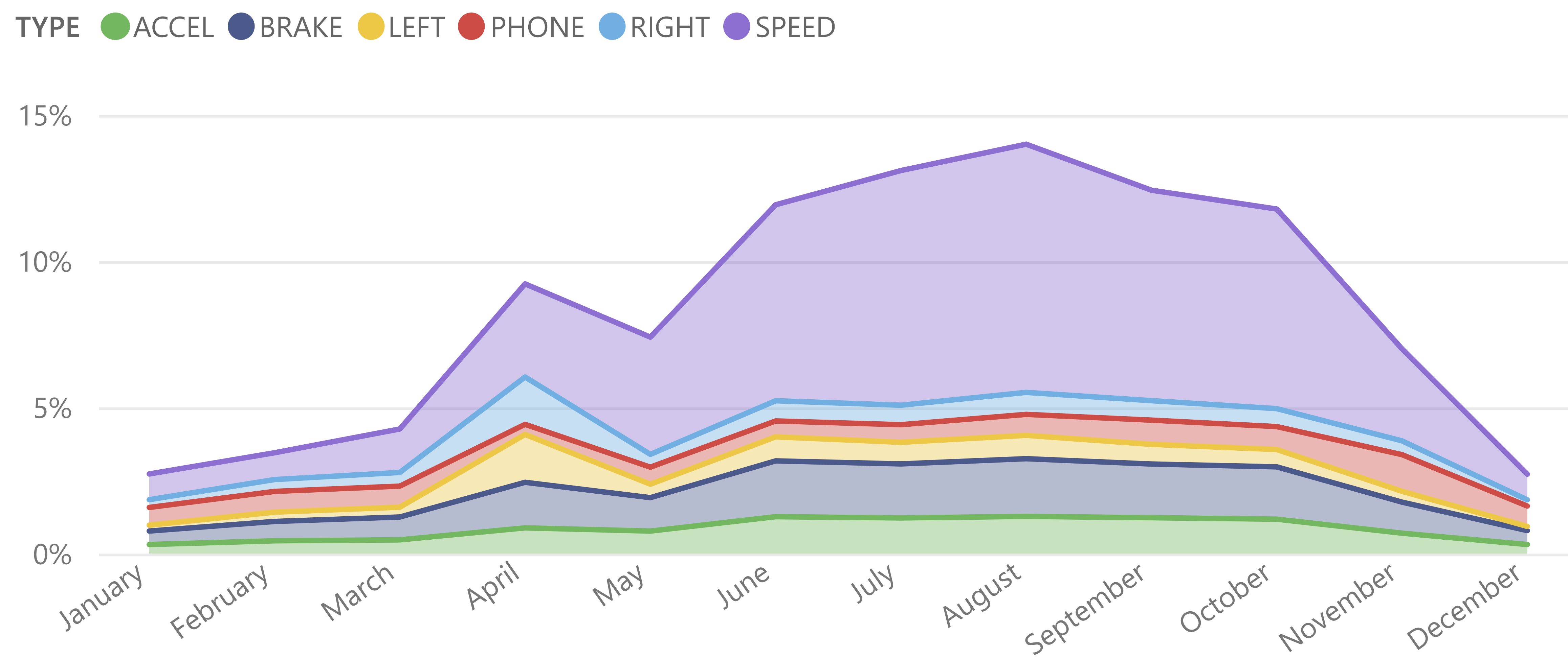
Trips Count

120.00K



Events Count

179.49K



Trips EDA

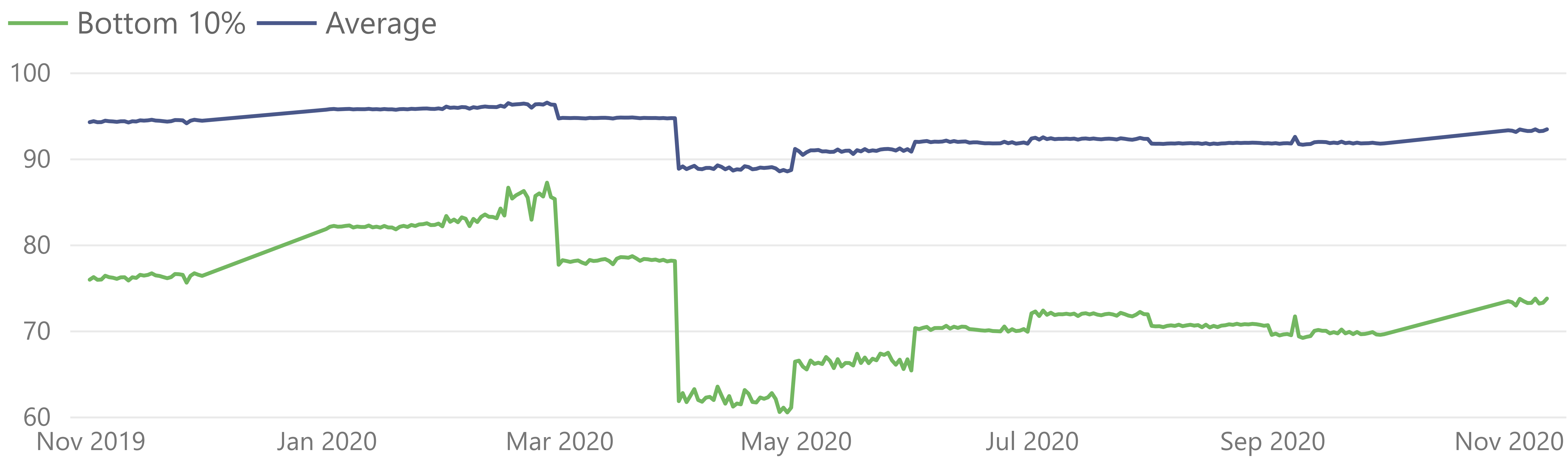
Start City

- [ ] Denver
- [ ] San Diego
- [ ] Seattle

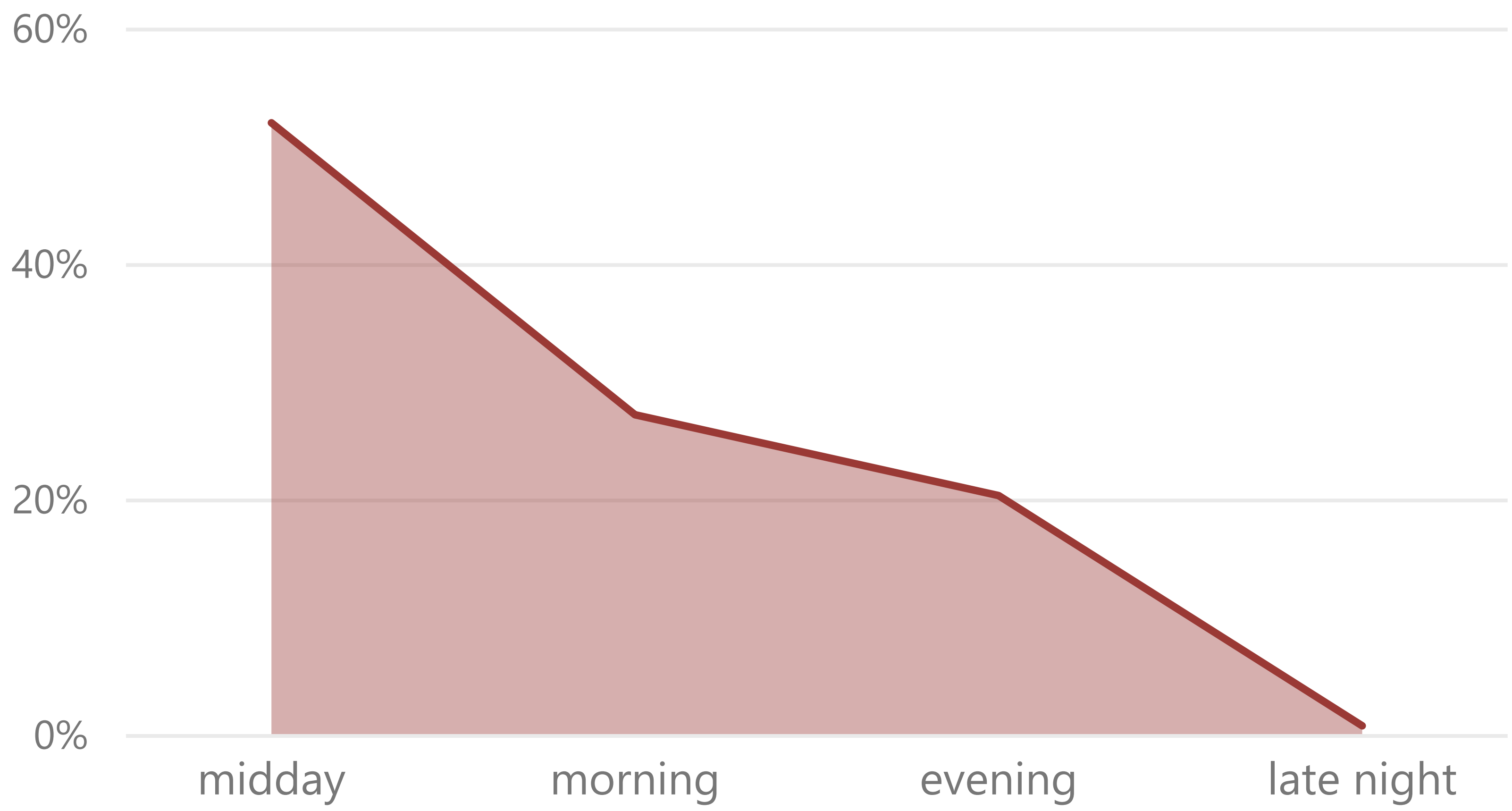
120.00K

Trip Count

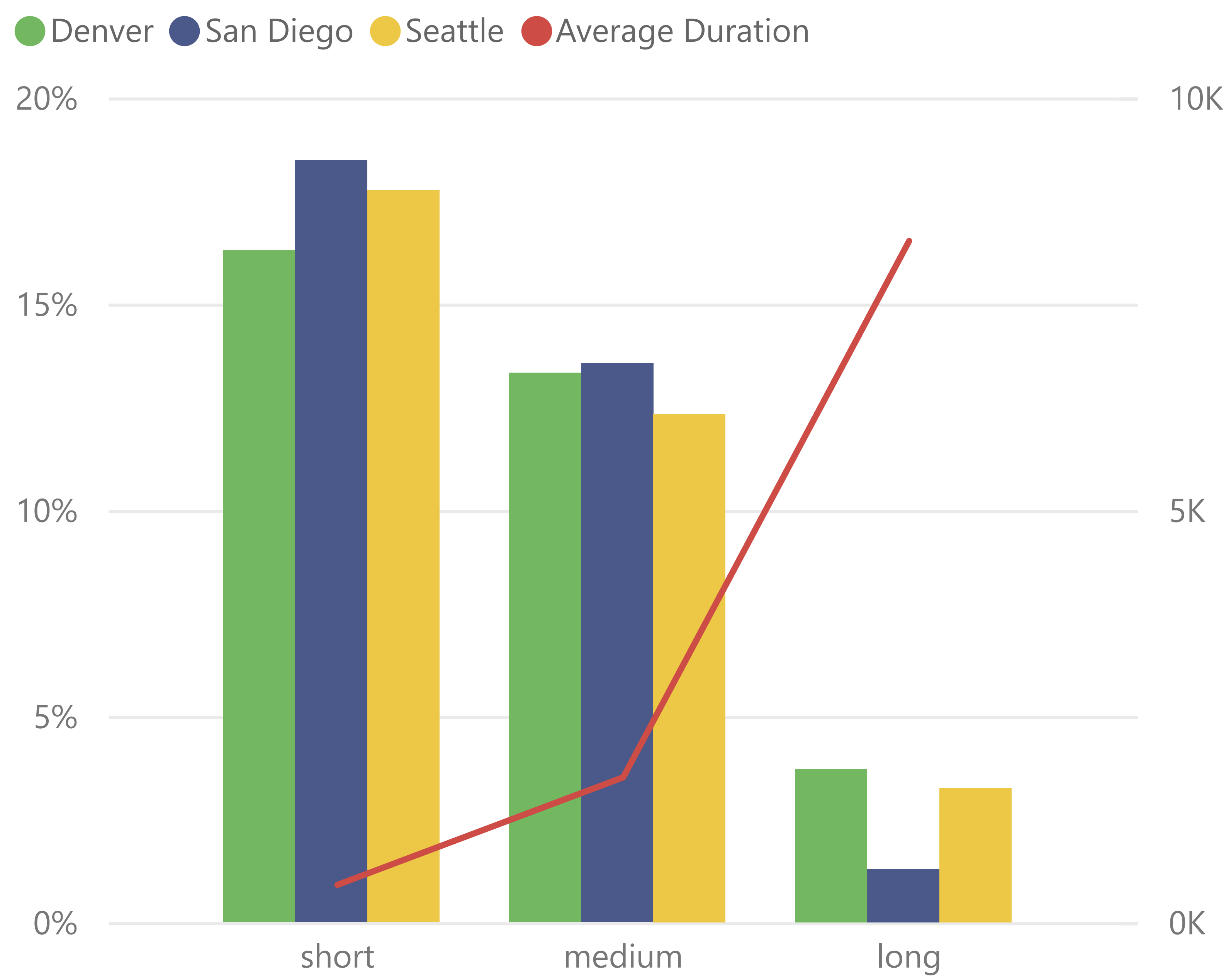
Speeding Scores Over Time



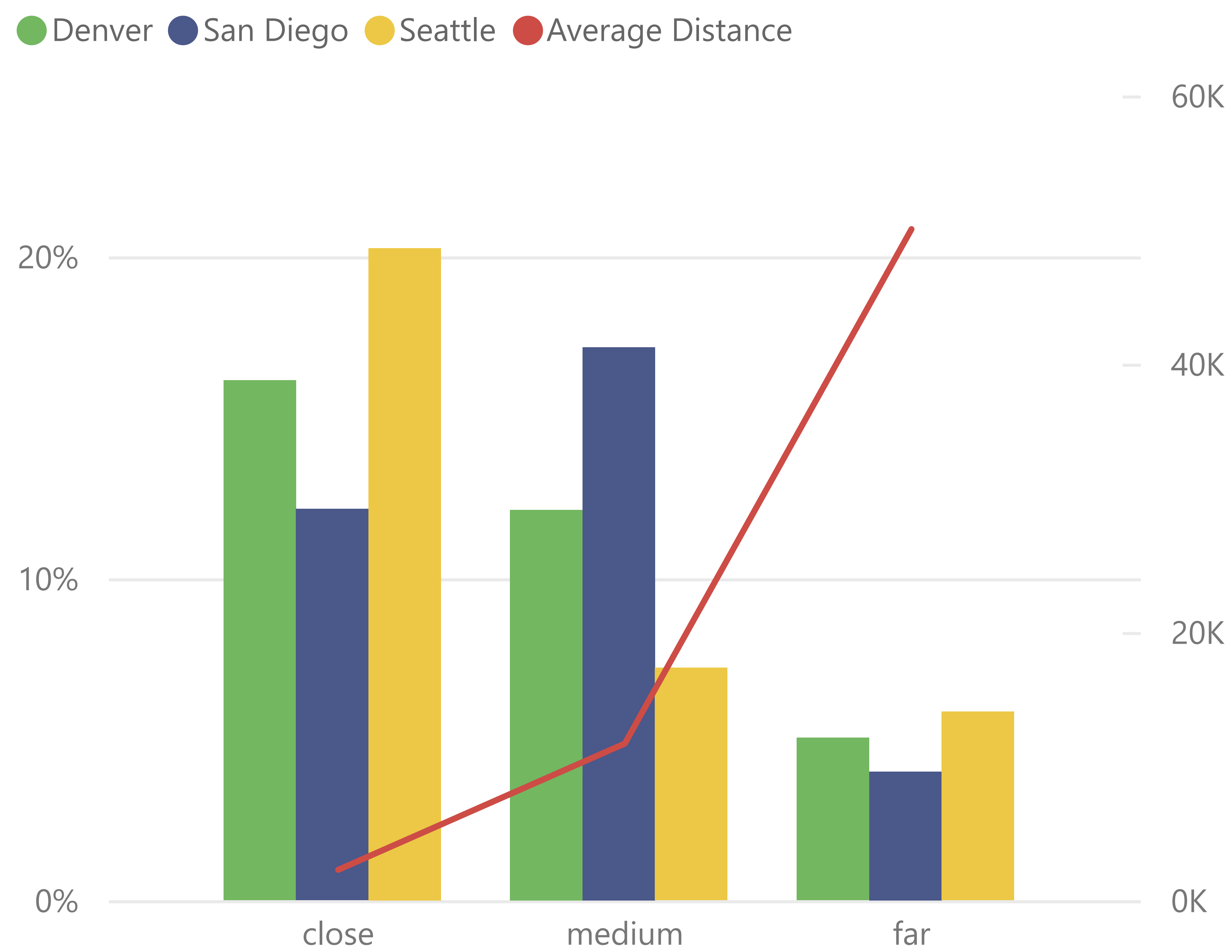
Trip Count by Time



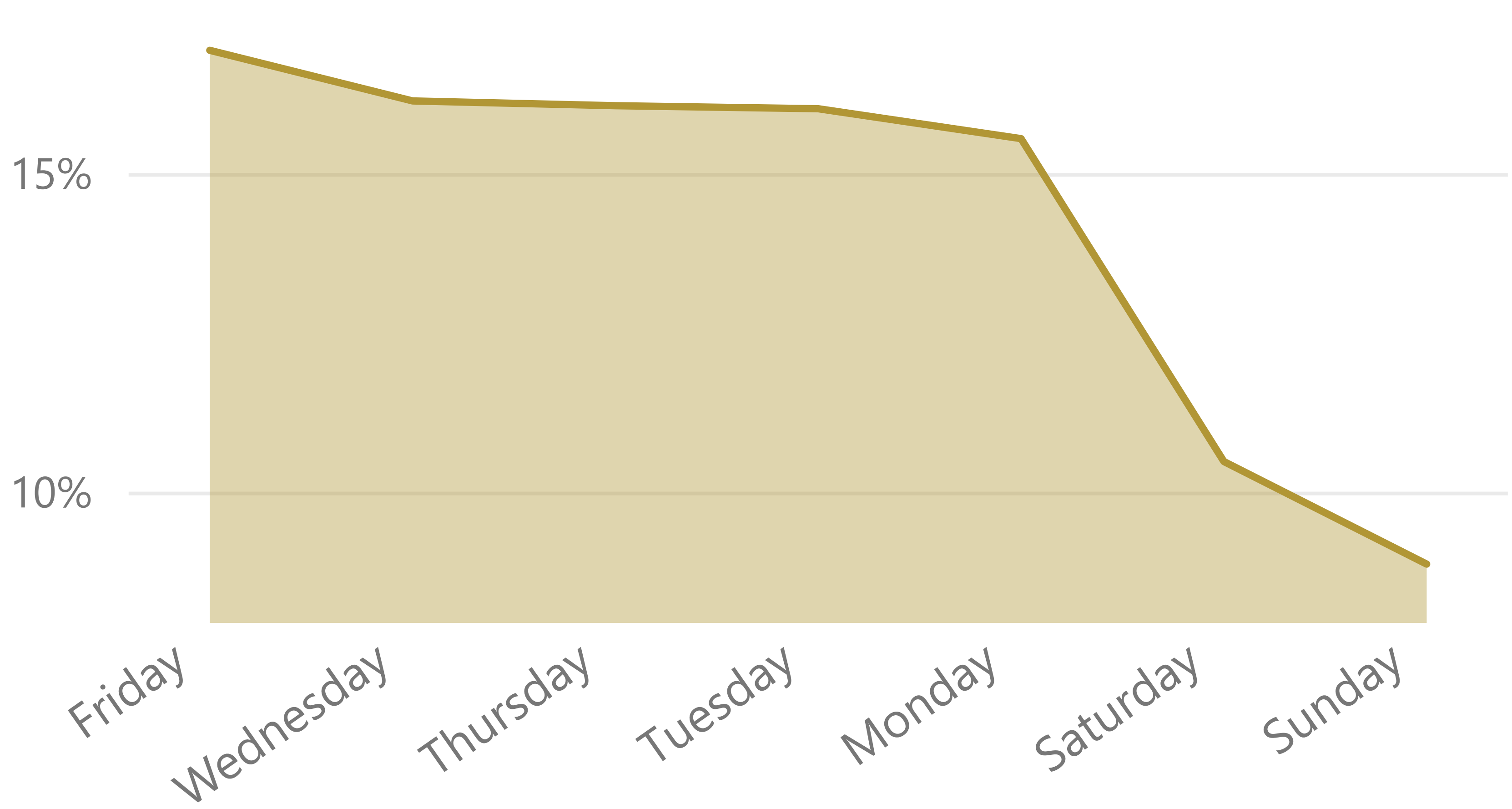
Trip Count by Duration



Trip Count by Distance



Trip Count by Day

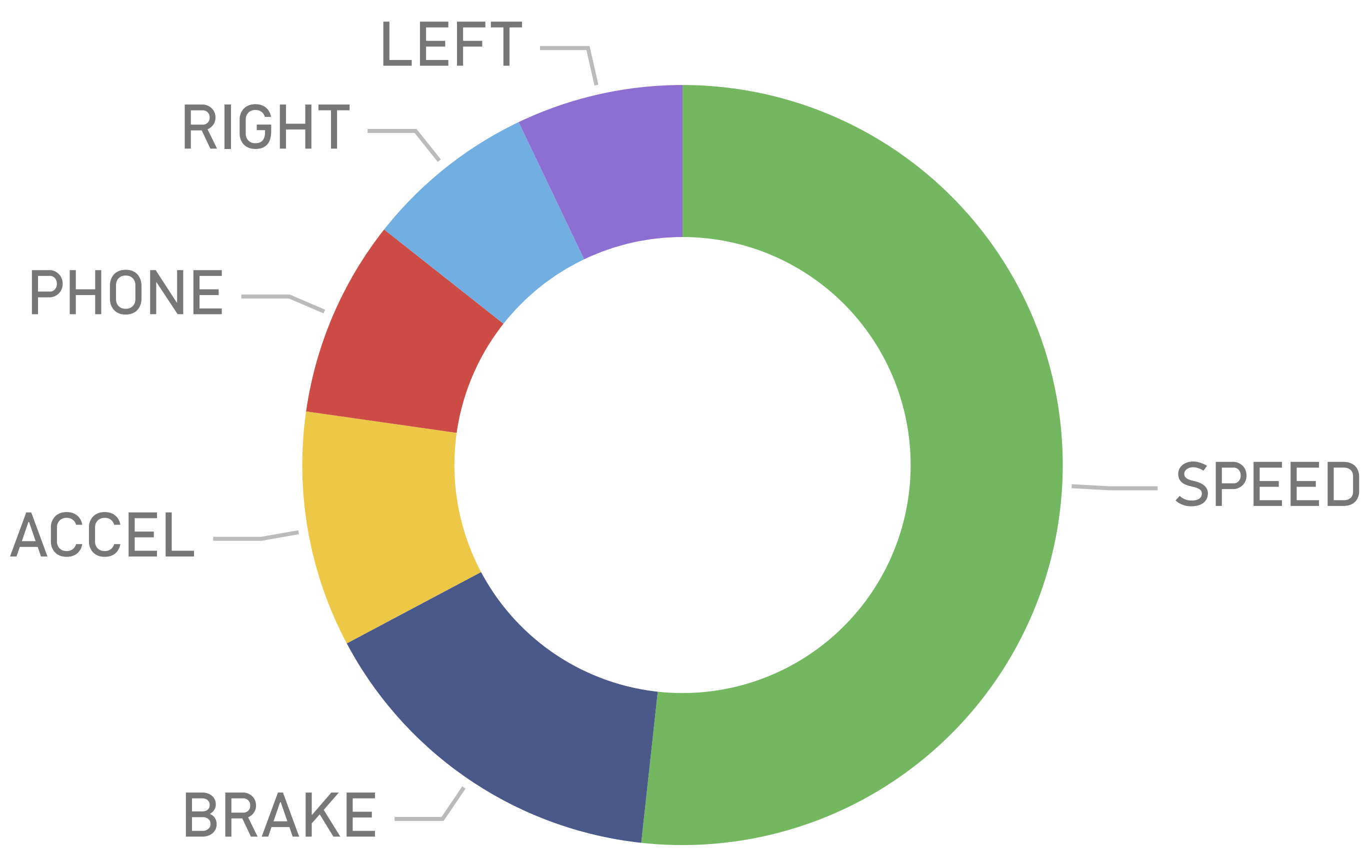


Telematics/Events EDA

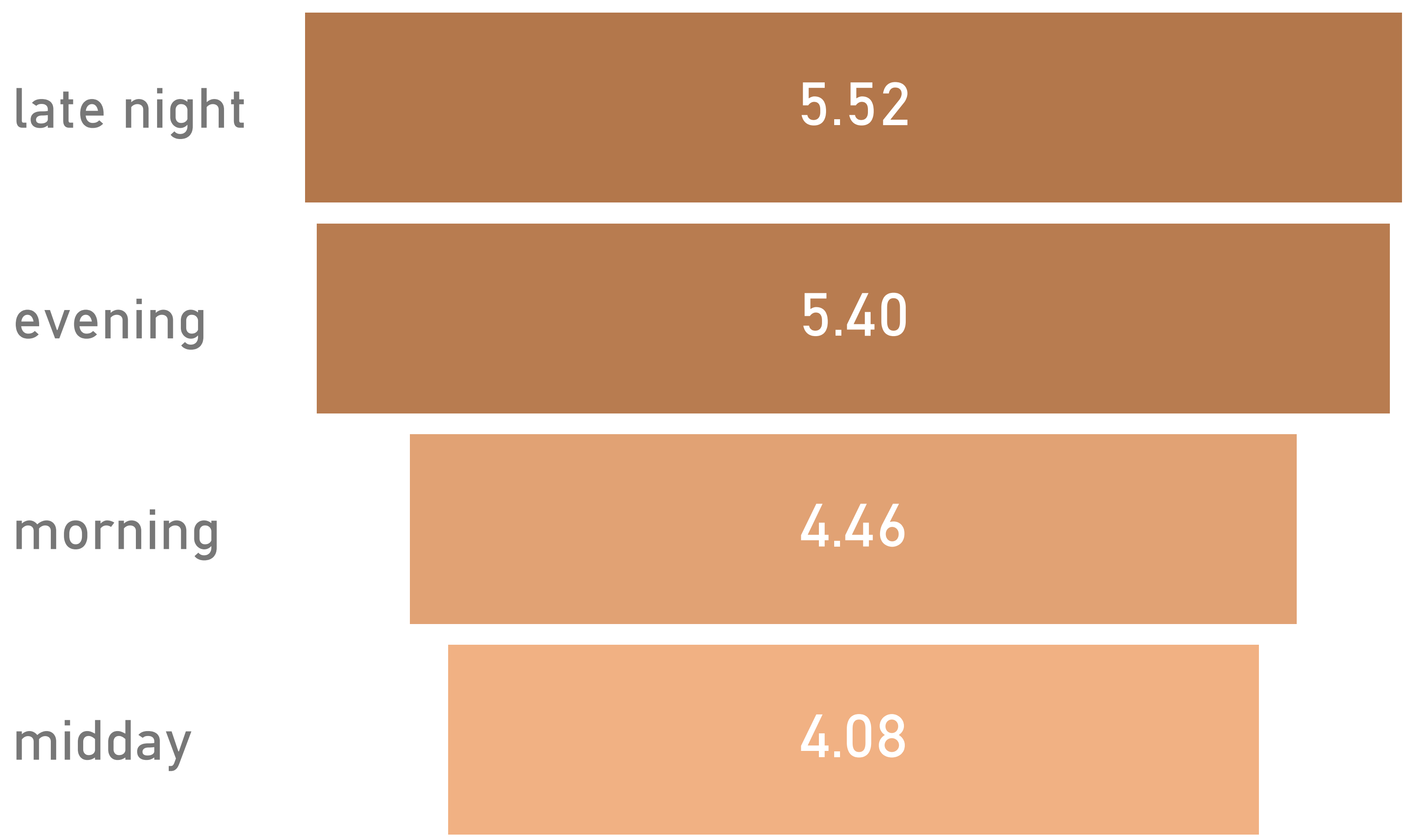
City

All

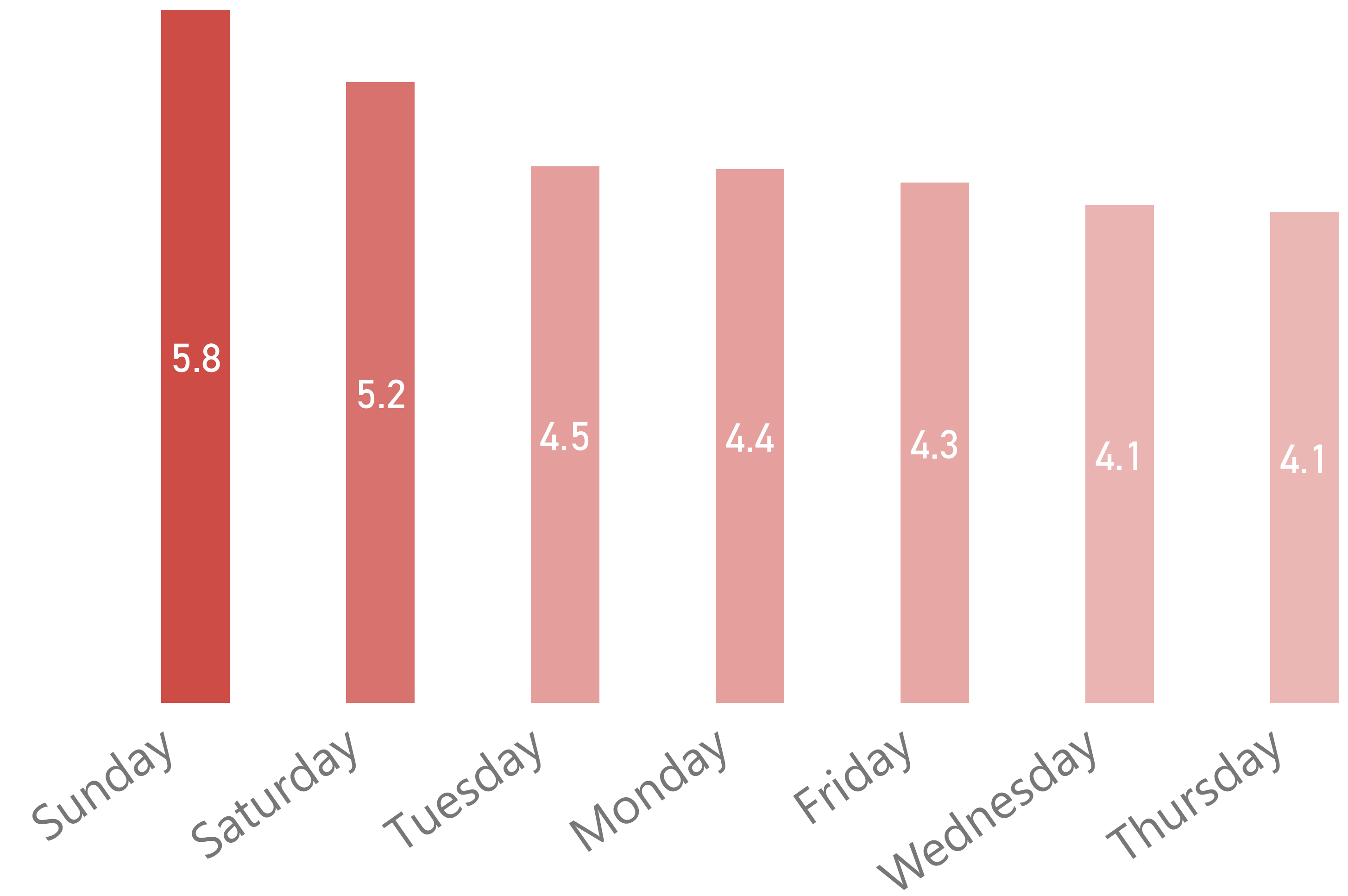
Events Types



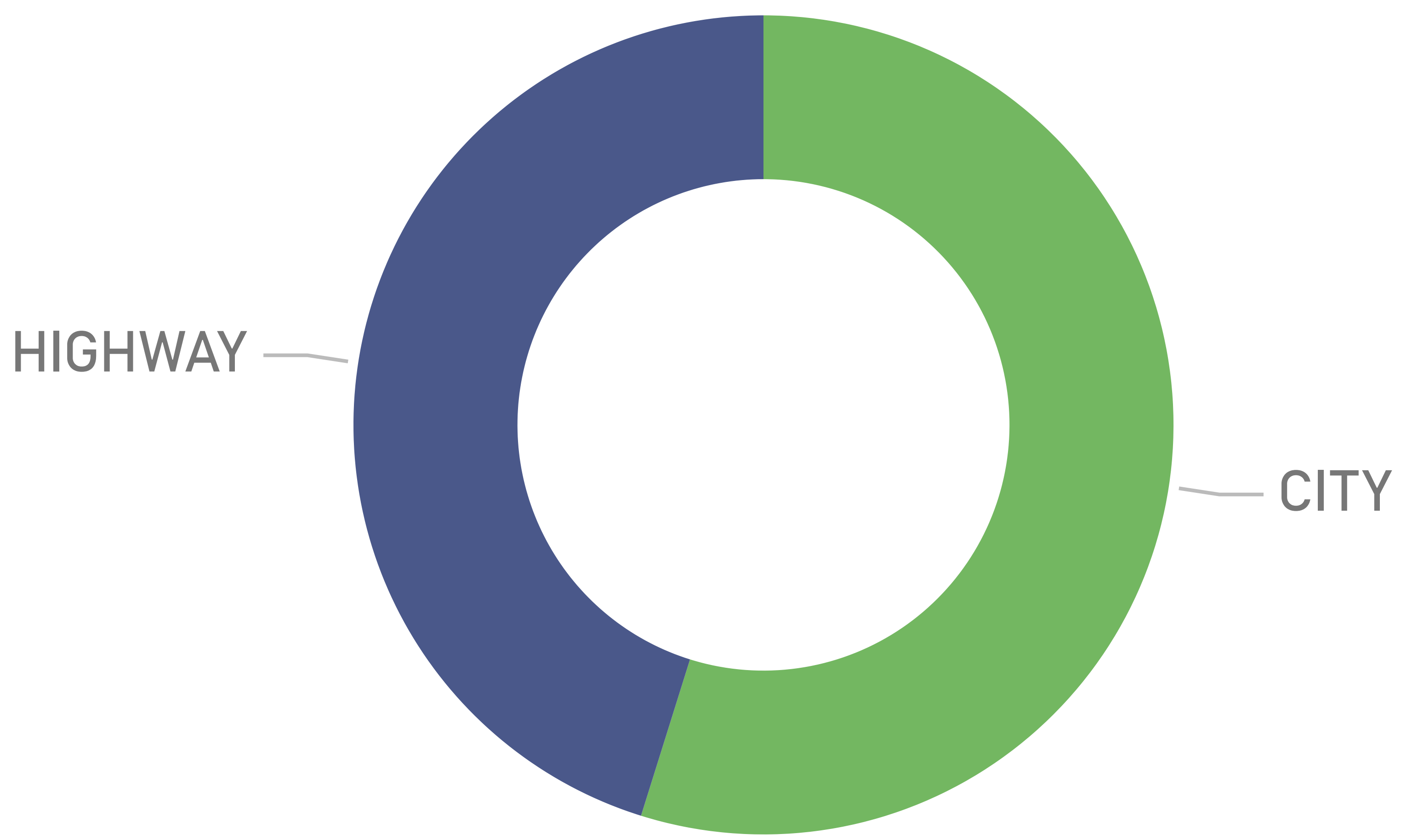
Ratio of Events by Time of Day



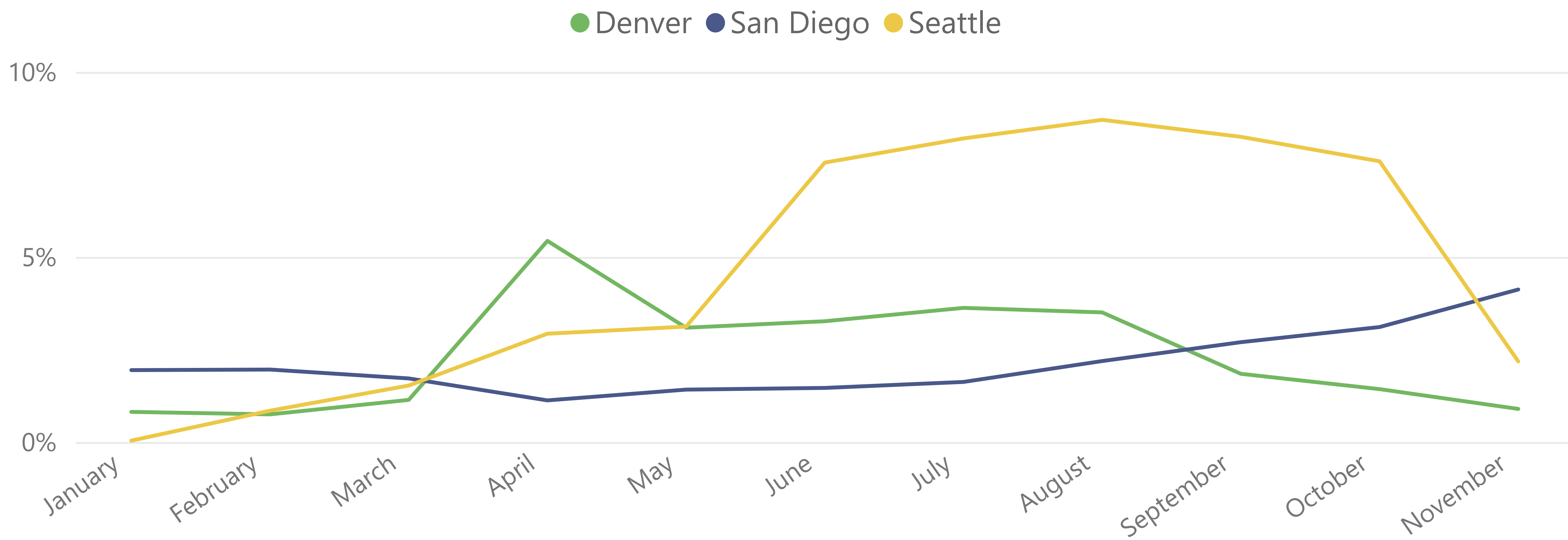
Ratio of Events by Day of the Week



Location of Events



Speeding Counts



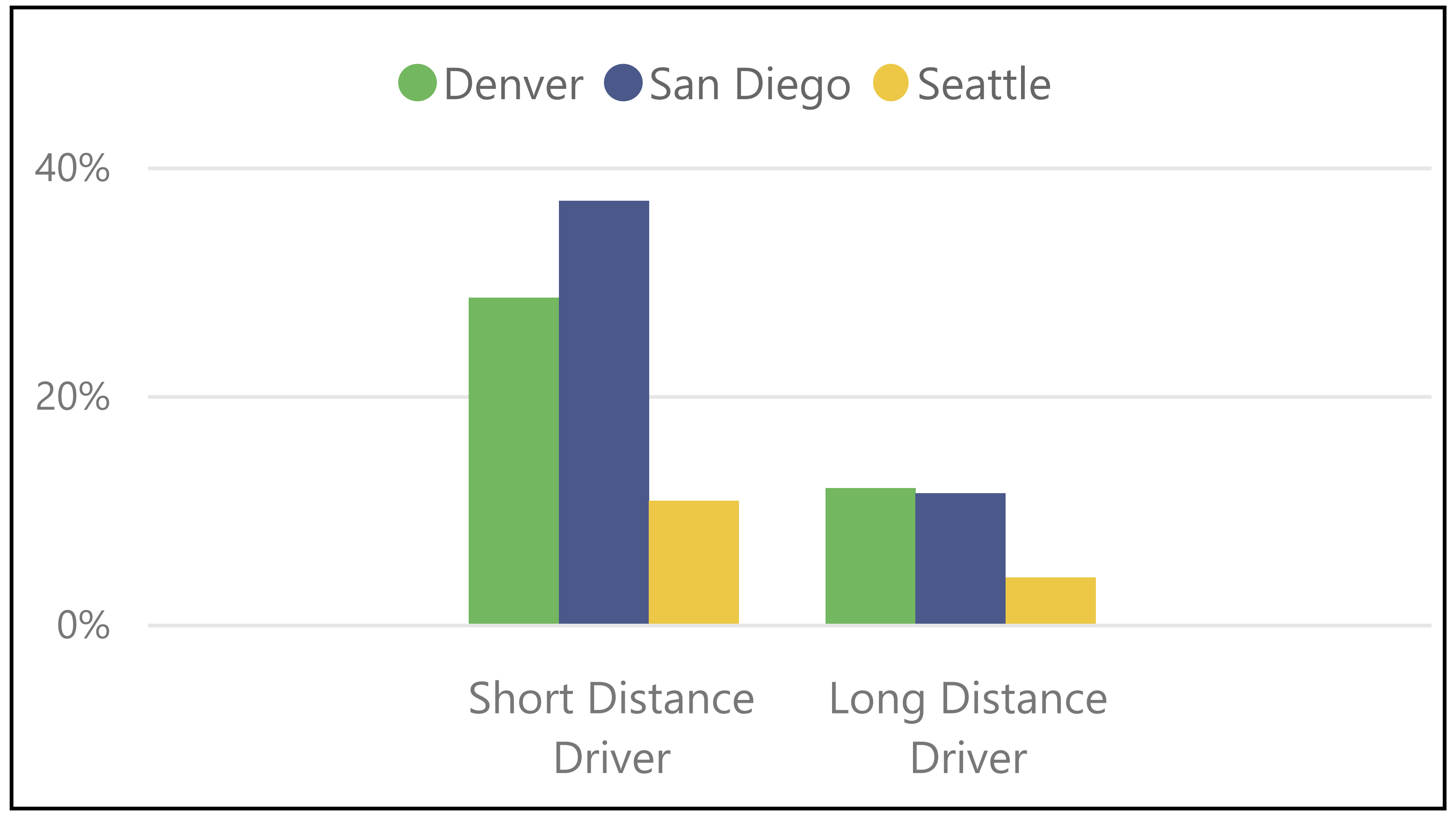
Exploring Driver Statistics

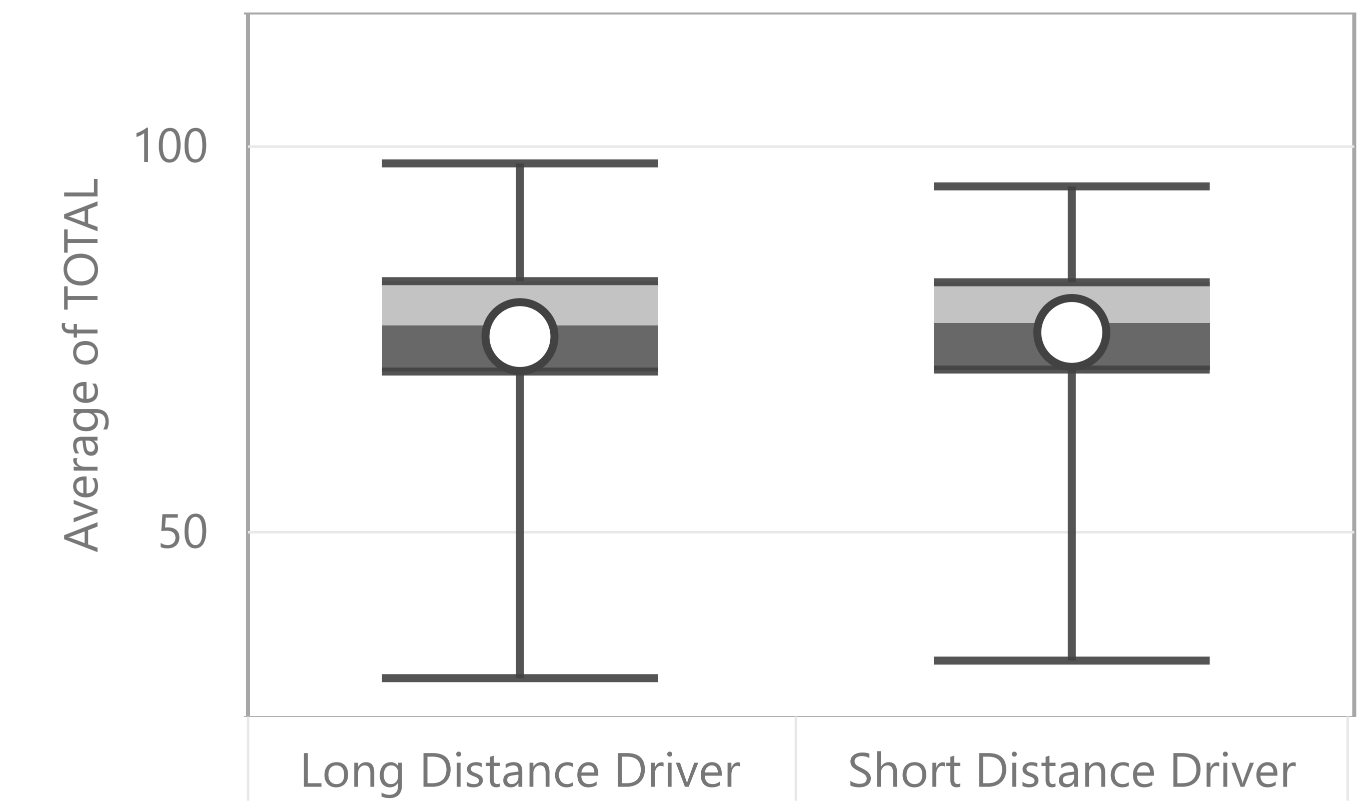
ID	Speeding	Distance Type	Days Type
9189	Normal	Short Distance Driver	Weekends Only
650778	High	Short Distance Driver	Weekends Only
652364	Normal	Short Distance Driver	Weekends Only
652532	Normal	Short Distance Driver	Weekends Only
652733	Normal	Short Distance Driver	Weekends Only
711548	Normal	Short Distance Driver	Weekends Only
713708	Normal	Short Distance Driver	Weekends Only
721319	High	Short Distance Driver	Weekends Only
721597	High	Short Distance Driver	Weekends Only
723929	High	Short Distance Driver	Weekends Only
723977	High	Short Distance Driver	Weekends Only
724825	Normal	Long Distance Driver	Weekends Only
725167	Normal	Short Distance Driver	Weekends Only
725619	Normal	Short Distance Driver	Weekends Only
727027	High	Long Distance Driver	Weekends Only
743393	Normal	Short Distance Driver	Weekends Only
744042	Normal	Short Distance Driver	Weekends Only
745080	Normal	Long Distance Driver	Weekends Only
745098	Normal	Short Distance Driver	Weekends Only
745400	High	Long Distance Driver	Weekends Only
747888	Normal	Short Distance Driver	Weekends Only
748501	Normal	Short Distance Driver	Weekends Only

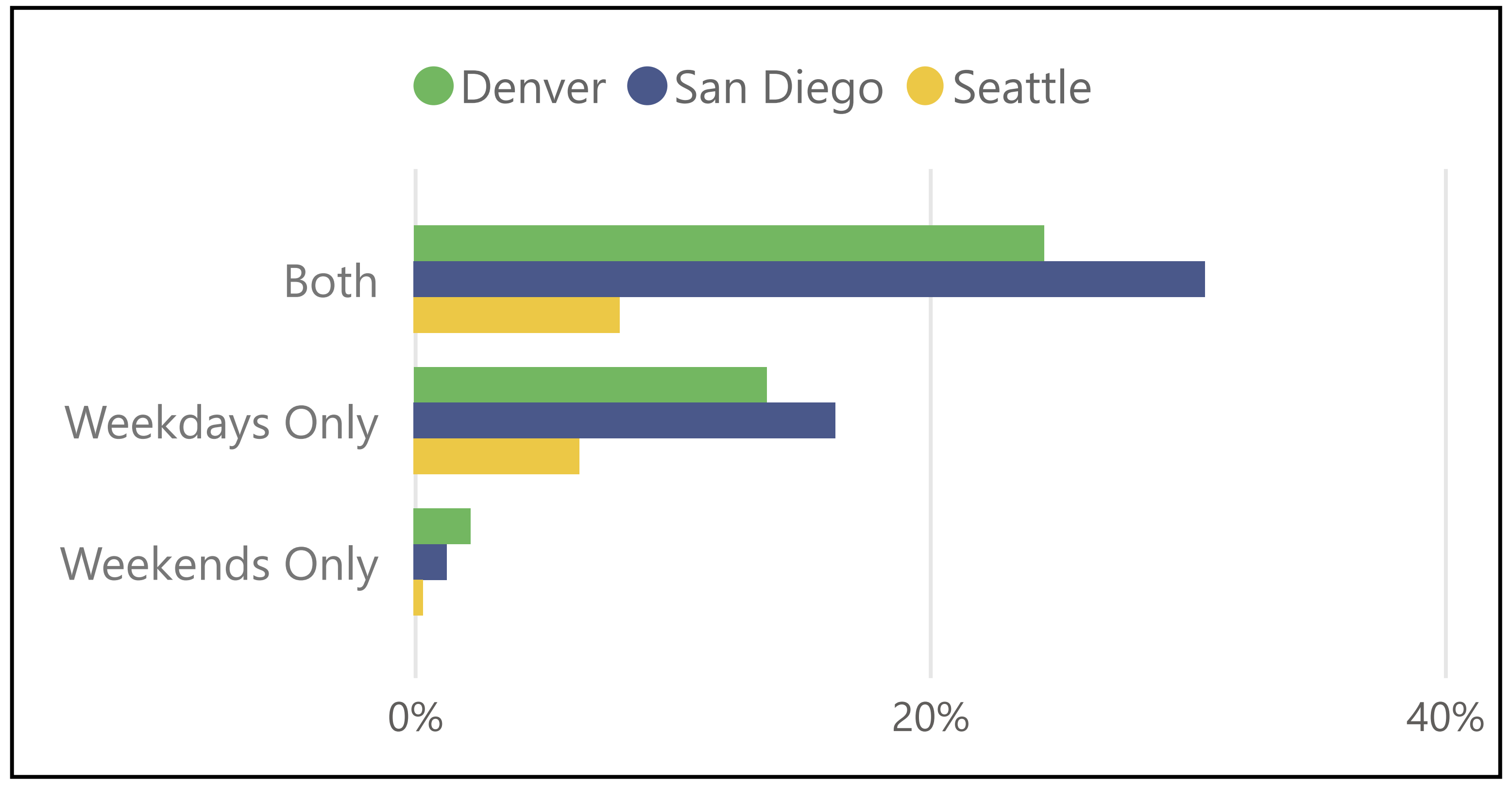
Driver Count

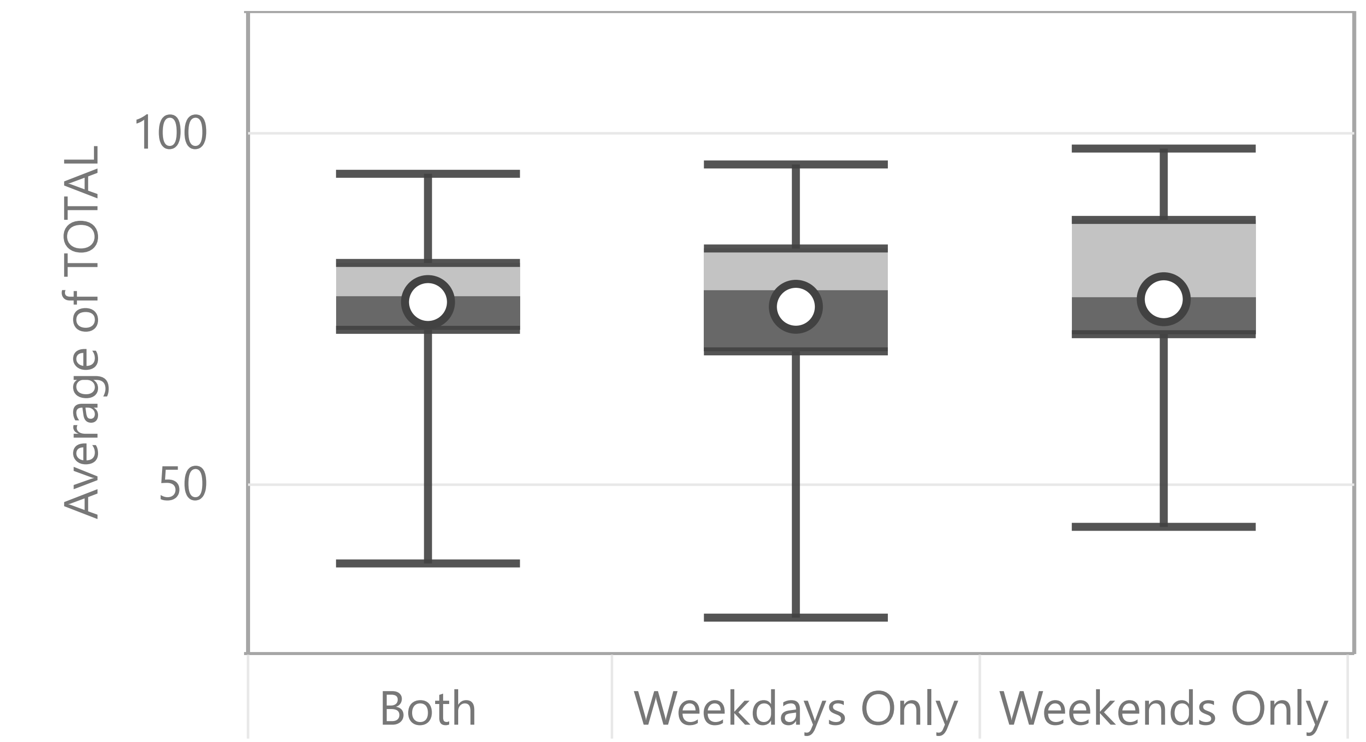
1193

- [x] Select all
- [x] Denver
- [x] San Diego
- [x] Seattle







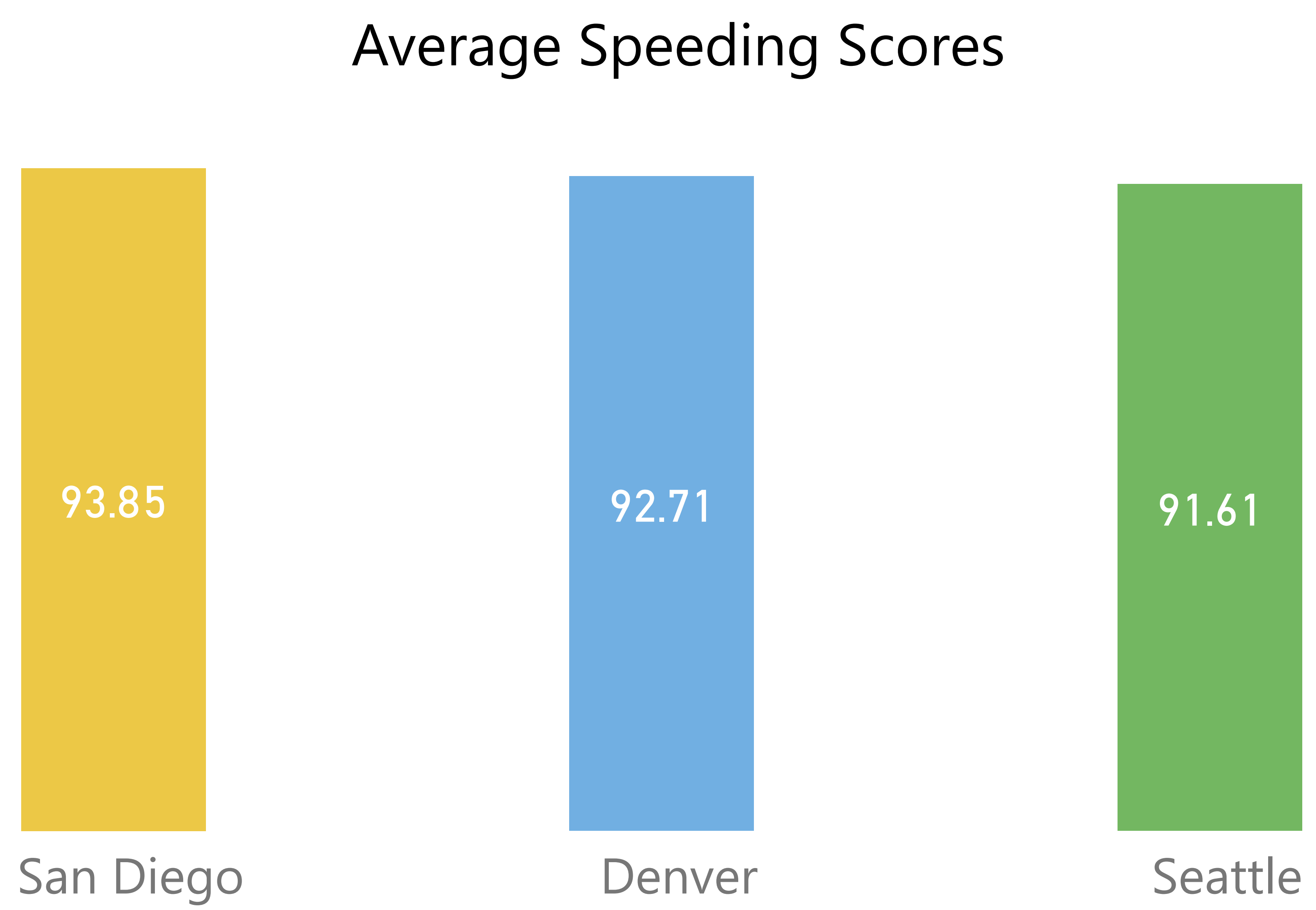


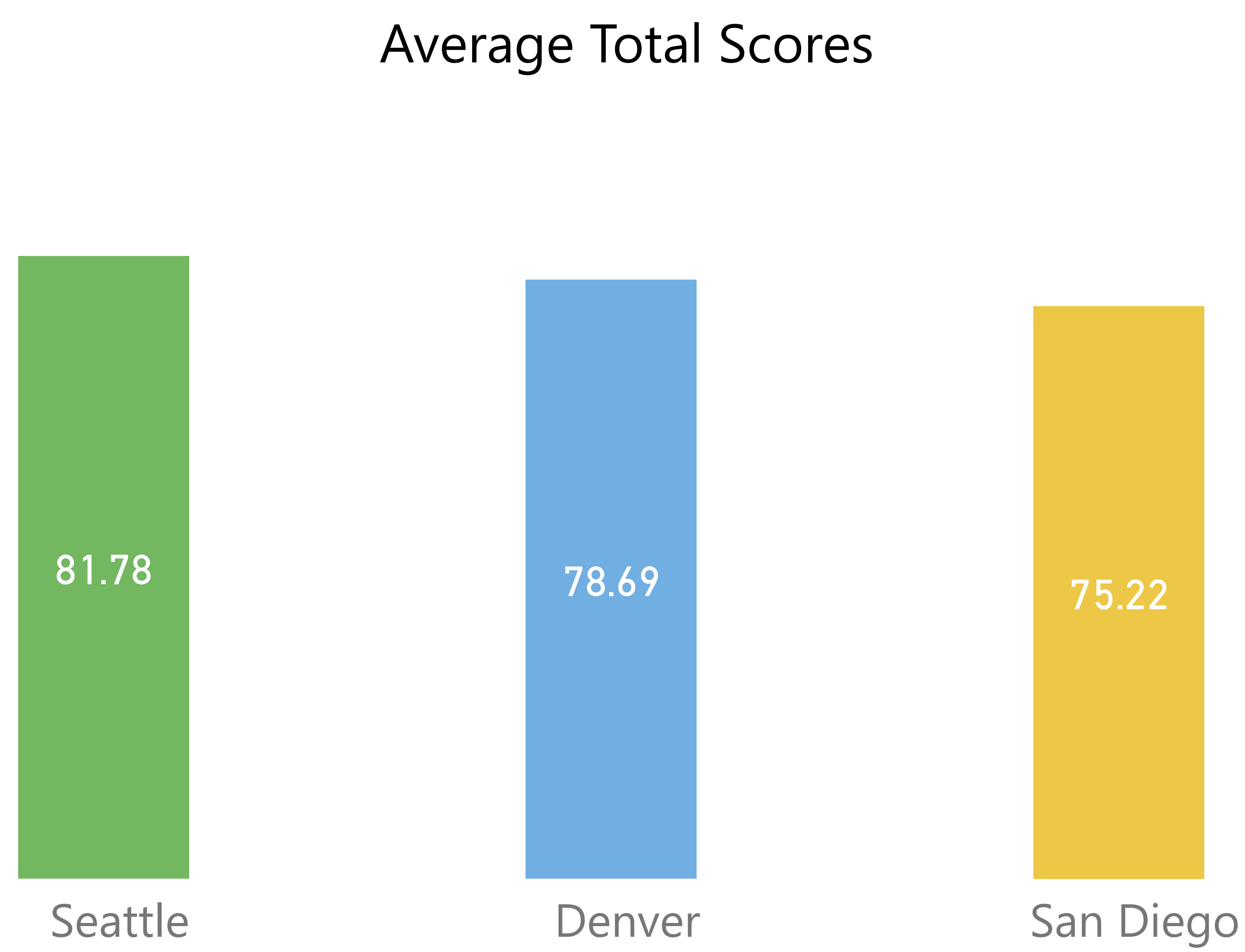
Statistical Analysis on Total and Speeding Scores

Type of Test:
Analysis of Variance (ANOVA) for the mean Total and Speeding scores among the different days of the week.

Null Hypothesis: There is no difference in mean total/mean speeding score among San Diego, Seattle, and Denver.

Alternative Hypothesis: There is at least one difference in mean total/speeding score between San Diego, Seattle, and Denver.





Results

Total Scores Between Cities

```
 Multiple Comparison of Means - Tukey HSD, FWER=0.05
=========================================================
 group1    group2  meandiff p-adj  lower   upper  reject
---------------------------------------------------------
   Denver San Diego   1.3151  0.001  0.5564  2.0737   True
   Denver   Seattle  -0.4485 0.3493 -1.2071  0.3102  False
San Diego   Seattle  -1.7635  0.001 -2.5222 -1.0049   True
---------------------------------------------------------
```

Speeding Scores Between Cities

```
 Multiple Comparison of Means - Tukey HSD, FWER=0.05
==========================================================
 group1    group2  meandiff p-adj  lower   upper  reject
----------------------------------------------------------
   Denver San Diego  -3.4357 0.001 -4.3056 -2.5659   True
   Denver   Seattle   3.4883 0.001  2.6184  4.3581   True
San Diego   Seattle    6.924 0.001  6.0542  7.7939   True
----------------------------------------------------------
```

Takeaway
Location of trips can affect the performance of drivers

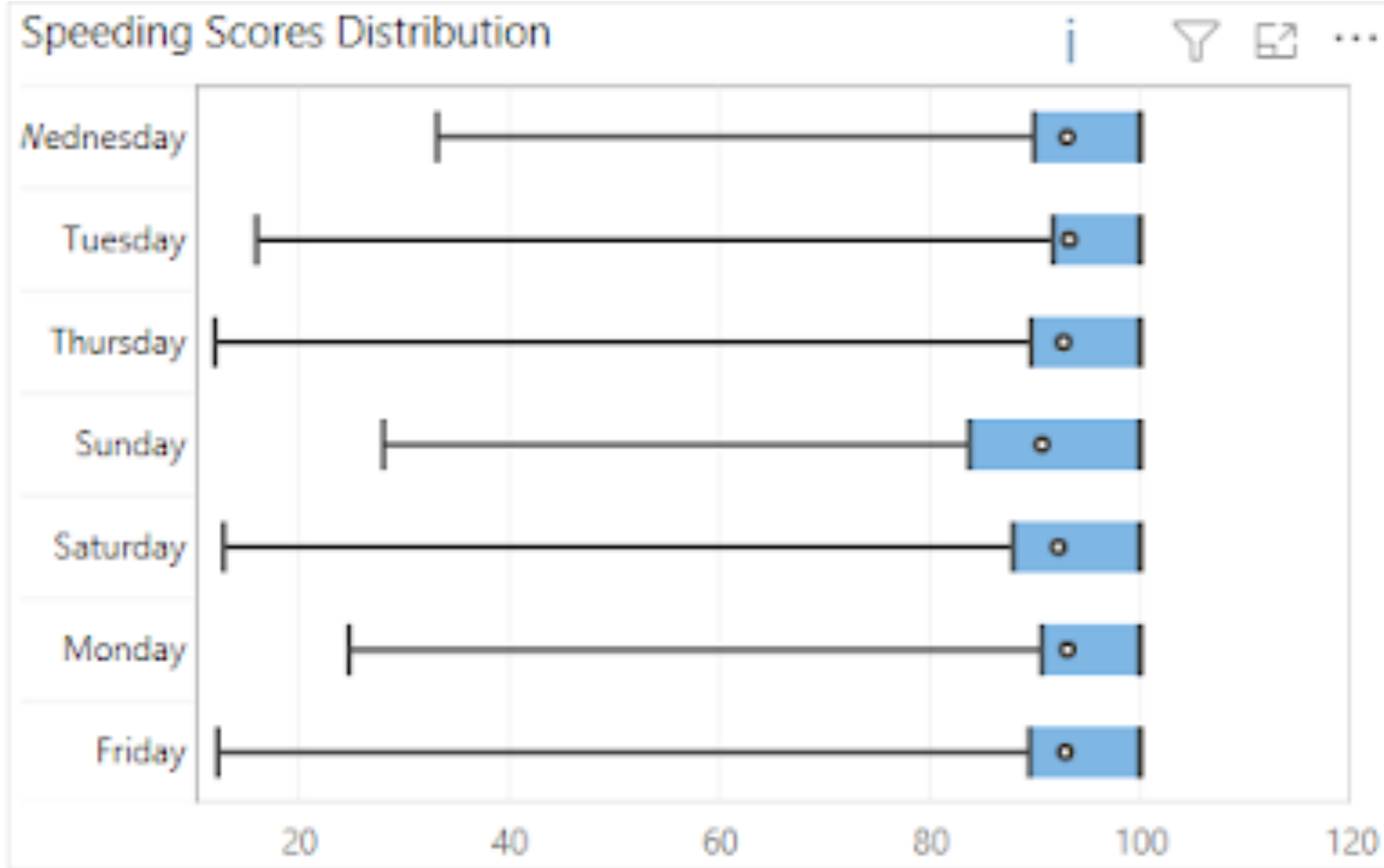
If the companies have service in both areas then they should have different ways to track their KPIs

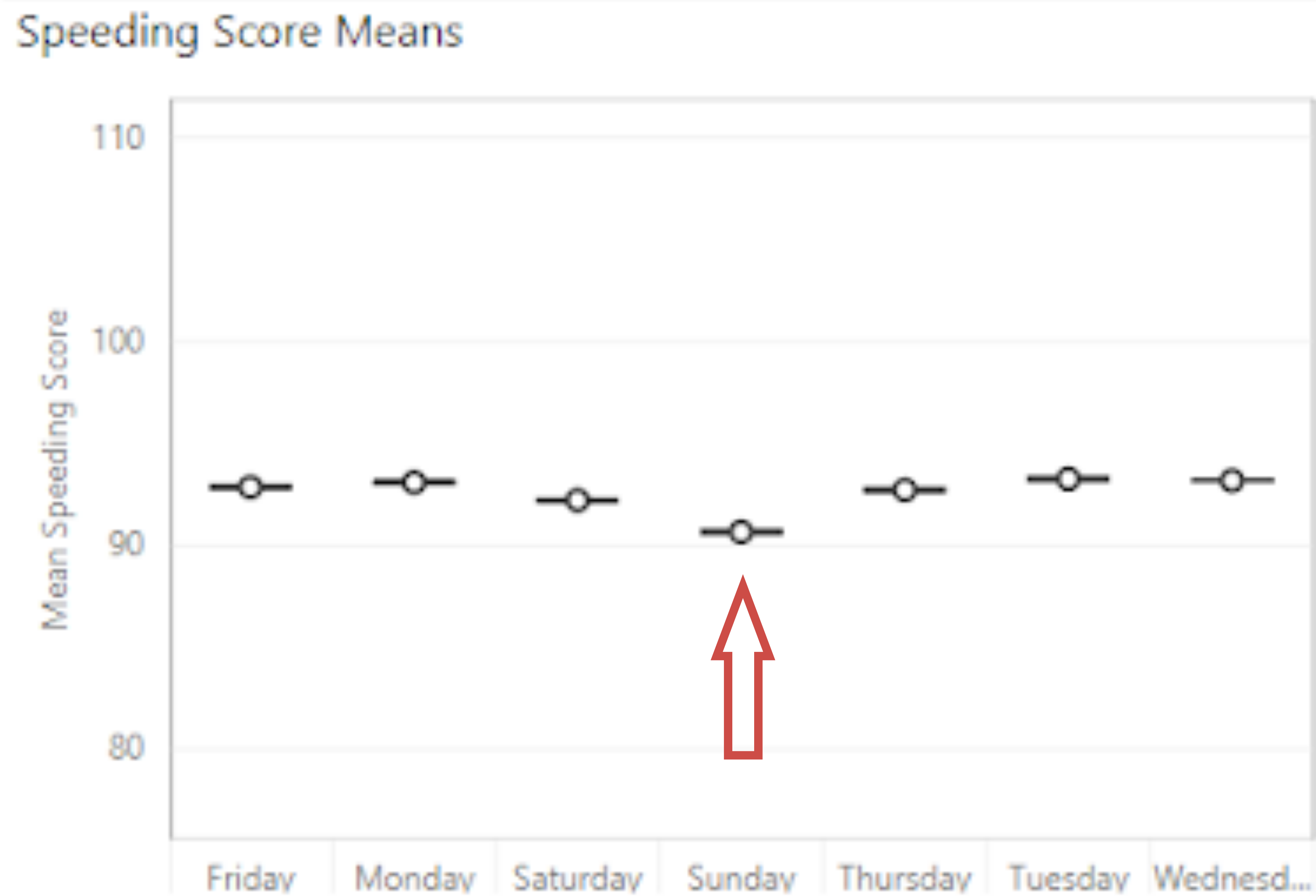
Speeding Scores Between Day of the Week

Type of Test:
Multiple paired comparison test for the mean Speeding scores among the different days of the week.

Null Hypothesis: There is <u>no difference in mean speeding score</u> among the two different days.

Alternative Hypothesis: There is <u>a difference in mean speeding score</u> between the two days.





Results

Multiple Comparison Test Between the Days of The Week

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Friday	Monday	0.3493	0.9	-2.6452	3.3438	False
Friday	Saturday	-1.3625	0.8099	-4.357	1.632	False
Friday	Sunday	-4.0143	0.0015	-7.0088	-1.0198	True
Friday	Thursday	-0.7893	0.9	-3.7838	2.2052	False
Friday	Tuesday	0.495	0.9	-2.4995	3.4895	False
Friday	Wednesday	0.2997	0.9	-2.6948	3.2942	False
Monday	Saturday	-1.7118	0.6077	-4.7063	1.2827	False
Monday	Sunday	-4.3636	0.001	-7.3581	-1.3691	True
Monday	Thursday	-1.1386	0.9	-4.1331	1.8559	False
Monday	Tuesday	0.1457	0.9	-2.8488	3.1402	False
Monday	Wednesday	-0.0496	0.9	-3.0441	2.9449	False
Saturday	Sunday	-2.6518	0.1227	-5.6463	0.3427	False
Saturday	Thursday	0.5732	0.9	-2.4213	3.5677	False
Saturday	Tuesday	1.8575	0.5234	-1.137	4.852	False
Saturday	Wednesday	1.6622	0.6364	-1.3323	4.6567	False
Sunday	Thursday	3.225	0.0252	0.2305	6.2195	True
Sunday	Tuesday	4.5093	0.001	1.5148	7.5038	True
Sunday	Wednesday	4.314	0.001	1.3195	7.3085	True
Thursday	Tuesday	1.2843	0.8552	-1.7102	4.2788	False
Thursday	Wednesday	1.089	0.9	-1.9055	4.0835	False
Tuesday	Wednesday	-0.1953	0.9	-3.1898	2.7992	False

Takeaway

Different days of the week can cause different Speeding performance -> Have different KPIs to track for different days of the week

Sunday's performance is different than the rest of the week -> Analyze "Sunday drivers"

Driver Scores KPIs & Benchmarking

ID

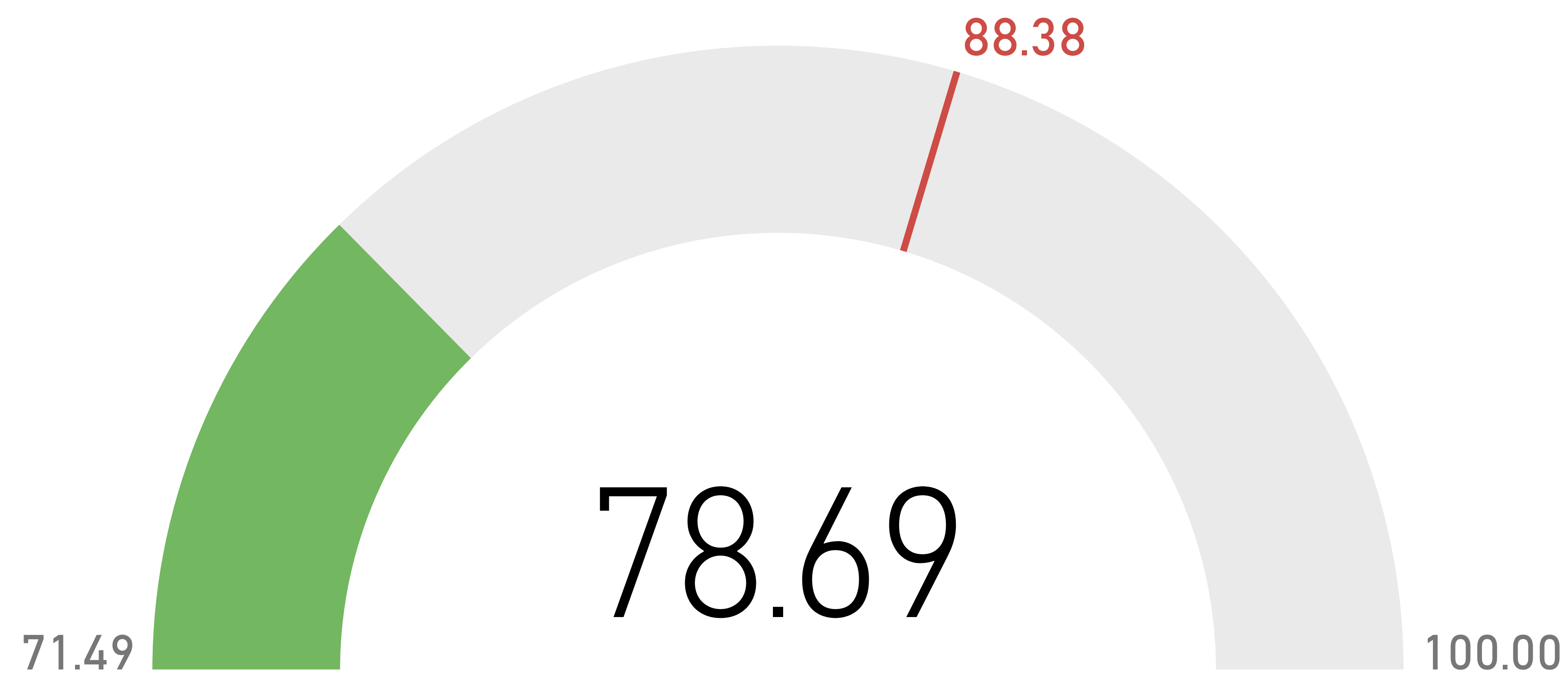
All

- [x] Denver
- [ ] San Diego
- [ ] Seattle

Day

- [ ] Friday
- [ ] Monday
- [ ] Saturday
- [ ] Sunday
- [ ] Thursday
- [ ] Tuesday
- [ ] Wednesday

Weighted Total Score



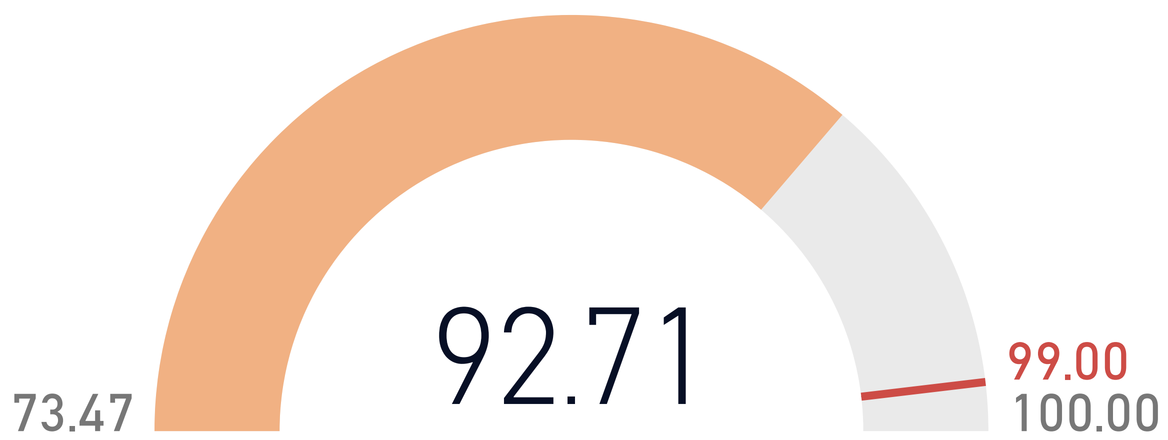
Month

- [x] Select all
- [x] April
- [x] August
- [x] December
- [x] February
- [x] January
- [x] July
- [x] June
- [x] March
- [x] May
- [x] November

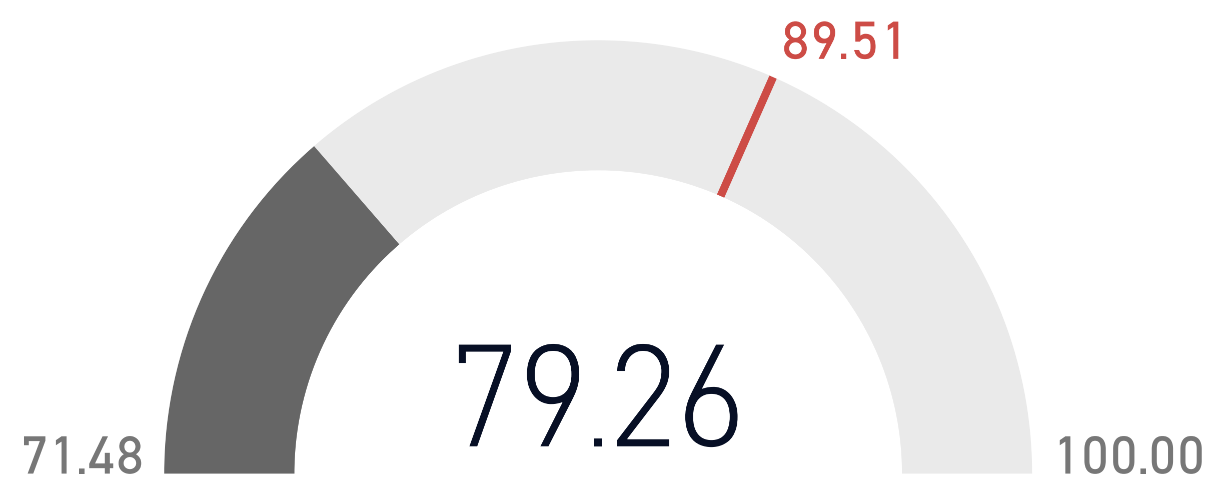
Trips Made

40.00K

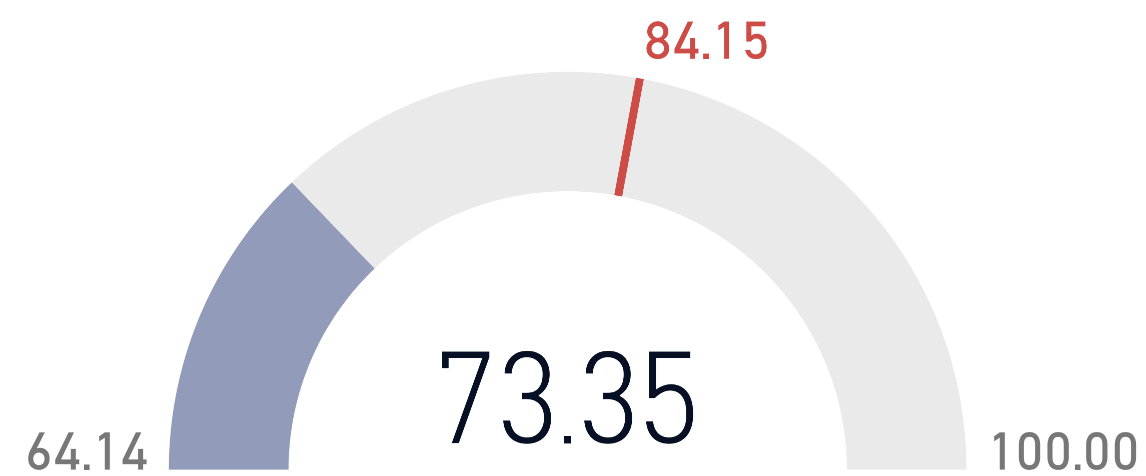
SPEEDING



SWERVE

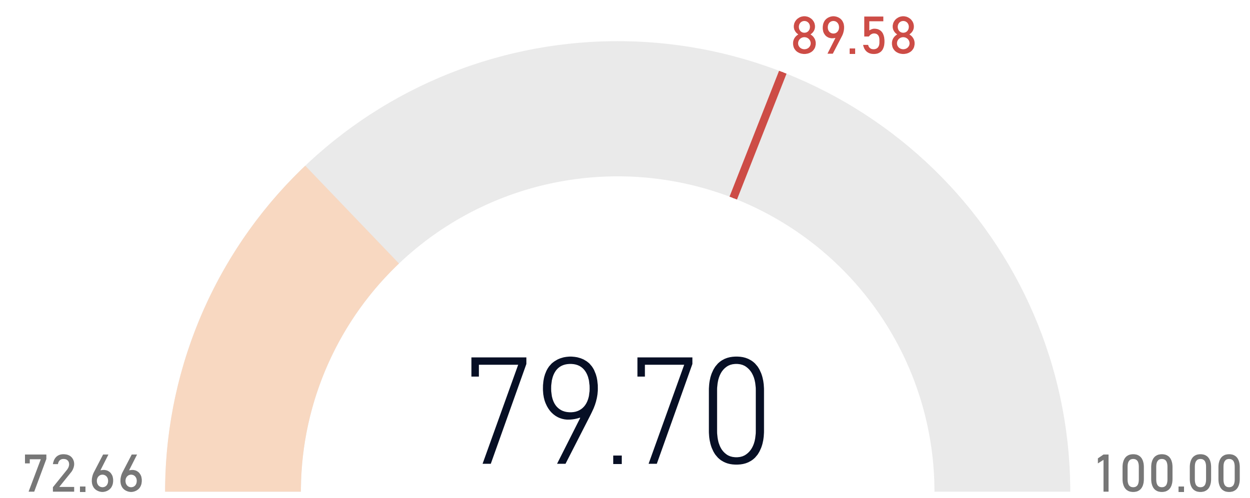


FUEL ECONOMY

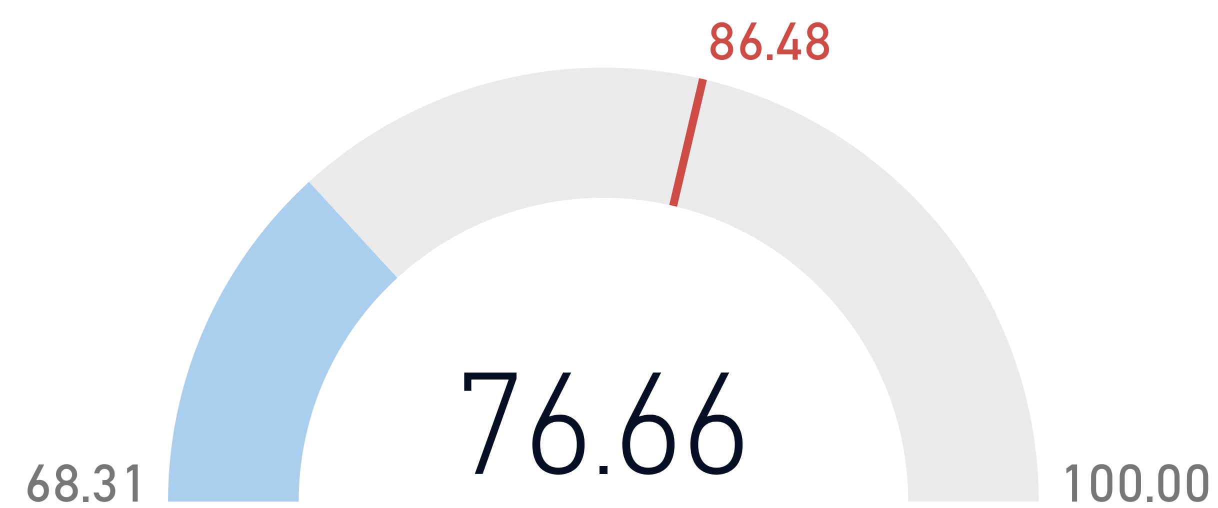


CHECKING PHONE

ALERTNESS



HANDLING / SMOOTHNESS



Conclusion

Future Implementations

• Plan to introduce external data sources

- Weather data
- COVID data to explain change in trips count

• Dive deeper into the speeding events, calculating new metrics, make better benchmarks -> analyze and prevent dangerous events

• Use machine learning (time-series models) to predict which drivers are going to speed or how much a driver will go over the speed limit