Visual story telling part 2 - Capital Metro data

September 14, 2023

```
[1]: import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns
```

```
[5]: cm=pd.read_csv('capmetro_UT.csv')
```

The daily hustle and bustle of university life is palpable, especially in how students and staff use the Capital Metro bus network around the UT Austin campus. Throughout the September to November 2018 semester, the following patterns emerged that paint a vivid picture of daily and monthly rhythms of bus ridership.

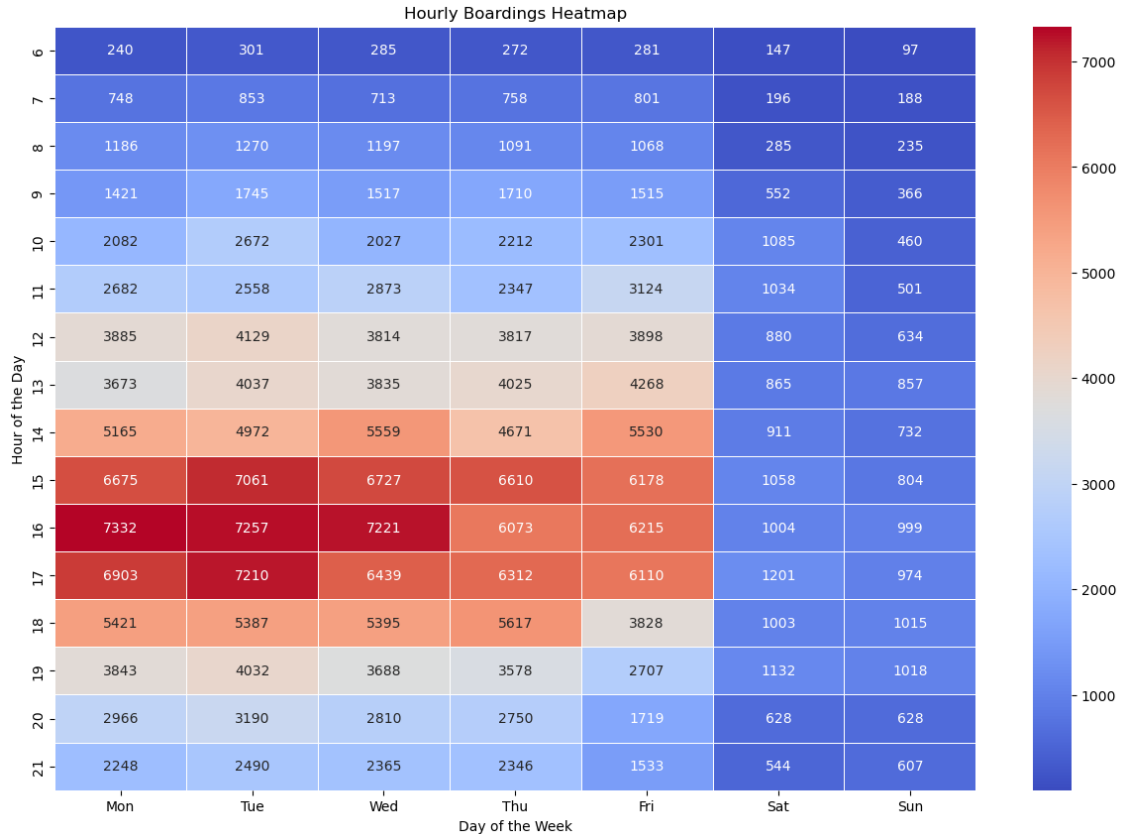## 0.1 1. The Weekly Rhythm: From Monday Blues to Weekend Respite

Hourly Boarding Heatmap : This showcases peak bus usage by hour and day, with color intensity indicating boarding . This snapshot helps identify high-demand periods for travel throughout the week.

```
[6]: # Group by hour_of_day and day_of_week for boarding
     grouped_boardings = cm.groupby(['hour_of_day', 'day_of_week']).boarding.sum().
      ↪unstack().fillna(0)

     # Weekdays ordering
     weekdays_order = ['Mon', 'Tue', 'Wed', 'Thu', 'Fri', 'Sat', 'Sun']

     grouped_boardings = grouped_boardings[weekdays_order]

     # Plot heatmap
     plt.figure(figsize=(15, 10))
     sns.heatmap(grouped_boardings, cmap="coolwarm", linewidths=0.5, annot=True,␣
      ↪fmt=".0f")
     plt.title("Hourly Boardings Heatmap")
     plt.ylabel("Hour of the Day")
     plt.xlabel("Day of the Week")
     plt.show()
```

Hourly Boardings Heatmap

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
6	240	301	285	272	281	147	97
7	748	853	713	758	801	196	188
8	1186	1270	1197	1091	1068	285	235
9	1421	1745	1517	1710	1515	552	366
10	2082	2672	2027	2212	2301	1085	460
11	2682	2558	2873	2347	3124	1034	501
12	3885	4129	3814	3817	3898	880	634
13	3673	4037	3835	4025	4268	865	857
14	5165	4972	5559	4671	5530	911	732
15	6675	7061	6727	6610	6178	1058	804
16	7332	7257	7221	6073	6215	1004	999
17	6903	7210	6439	6312	6110	1201	974
18	5421	5387	5395	5617	3828	1003	1015
19	3843	4032	3688	3578	2707	1132	1018
20	2966	3190	2810	2750	1719	628	628
21	2248	2490	2365	2346	1533	544	607

Hour of the Day / Day of the Week

Insights:

The Hourly Boarding Heatmap shows a clear pattern: from Monday to Thursday, the UT community's boarding peaks from 2 pm to 6 pm. By Friday, this trend shifts slightly earlier, with peak usage from 2 pm to 5 pm. This possibly indicates students and staff wrapping up their day earlier, eager for the weekend. And as one might anticipate, the weekends are significantly quieter, reflecting a campus at rest.
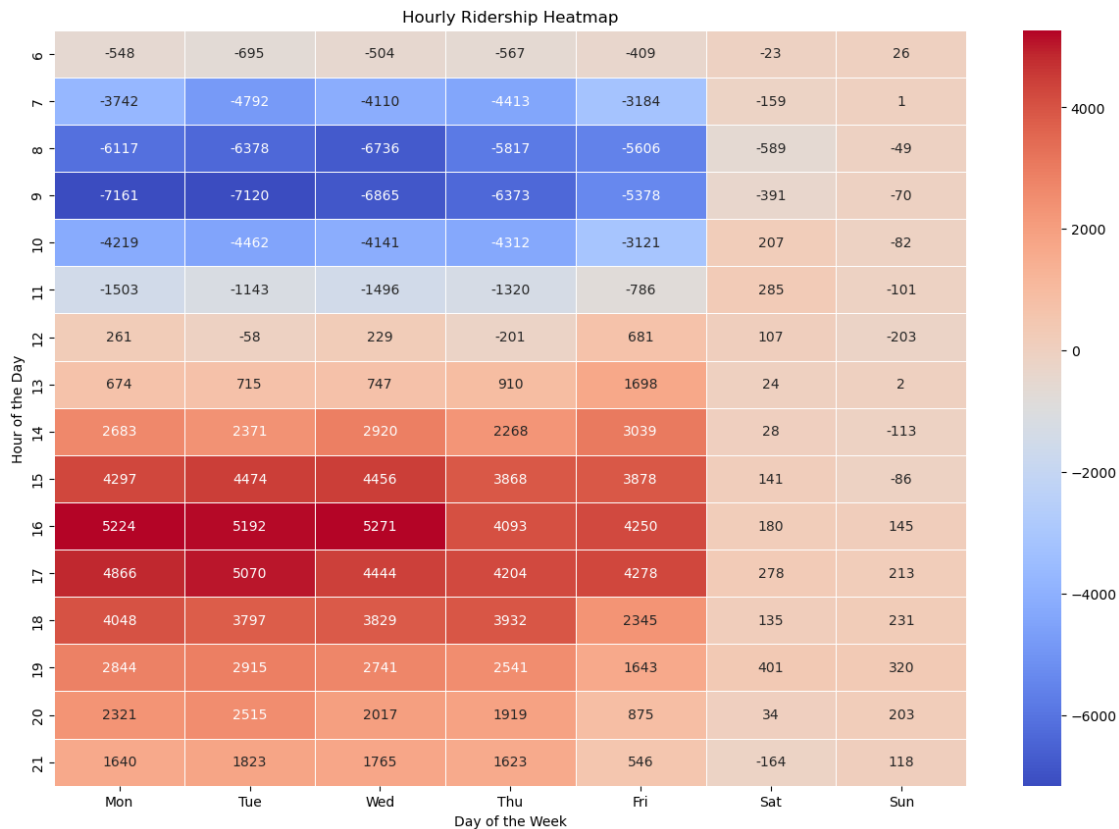
Hourly Ridership Heatmap : Similar to the above, but here, the color intensity indicates net ridership (boarding - alighting). Note that negative ridership figures in the context of our data indicate that more people alighted (got off) the bus than boarded (got on) during a given time window.

```
[7]: # Calculate net ridership
     cm['net_ridership'] = cm['boarding'] - cm['alighting']

     # Group by hour_of_day and day_of_week
     grouped_nr = cm.groupby(['hour_of_day', 'day_of_week']).net_ridership.sum().
     ↪unstack().fillna(0)


     grouped_nr = grouped_nr[weekdays_order]
```

```
# Plot heatmap
plt.figure(figsize=(15, 10))
sns.heatmap(grouped_nr, cmap="coolwarm", linewidths=0.5, annot=True, fmt=".0f")
plt.title("Hourly Ridership Heatmap")
plt.ylabel("Hour of the Day")
plt.xlabel("Day of the Week")
plt.show()
```

Hourly Ridership Heatmap

Hour of the Day	Mon	Tue	Wed	Thu	Fri	Sat	Sun
6	-548	-695	-504	-567	-409	-23	26
7	-3742	-4792	-4110	-4413	-3184	-159	1
8	-6117	-6378	-6736	-5817	-5606	-589	-49
9	-7161	-7120	-6865	-6373	-5378	-391	-70
10	-4219	-4462	-4141	-4312	-3121	207	-82
11	-1503	-1143	-1496	-1320	-786	285	-101
12	261	-58	229	-201	681	107	-203
13	674	715	747	910	1698	24	2
14	2683	2371	2920	2268	3039	28	-113
15	4297	4474	4456	3868	3878	141	-86
16	5224	5192	5271	4093	4250	180	145
17	4866	5070	4444	4204	4278	278	213
18	4048	3797	3829	3932	2345	135	231
19	2844	2915	2741	2541	1643	401	320
20	2321	2515	2017	1919	875	34	203
21	1640	1823	1765	1623	546	-164	118

Insights:

The Hourly Ridership Heatmap further deepens this understanding. With the net ridership metric (boarding - alighting), it's evident that from Monday to Thursday, the busiest bus periods are from 2 pm to 8 pm. By Friday, this shifts to 1 pm to 7 pm. Again, weekends display minimal activity, which aligns with our boarding insights.

0.2   2. Temperature's subtle influence

Average Boardings and Temperature by hour of the day :

[9]:

```python
# Group by hour_of_day
grouped_hr = cm.groupby('hour_of_day').agg(avg_temp=('temperature', 'mean'),
  ↪avg_boarding=('boarding', 'mean')).reset_index()

# Set up the figure and axes
fig, ax1 = plt.subplots(figsize=(12, 7))

# Plot average boardings on primary y-axis
ax1.bar(grouped_hr['hour_of_day'], grouped_hr['avg_boarding'], color='blue',
  ↪alpha=0.6, label='Average Boardings')
ax1.set_xlabel('Hour of the Day')
ax1.set_ylabel('Average Number of Boardings', color='blue')
ax1.tick_params('y', colors='blue')
ax1.set_xticks(grouped_hr['hour_of_day'])  # To ensure all hours are shown

# Create the secondary y-axis to plot average temperature
ax2 = ax1.twinx()
ax2.plot(grouped_hr['hour_of_day'], grouped_hr['avg_temp'], color='red',
  ↪marker='o', label='Average Temperature')
ax2.set_ylabel('Average Temperature (°F)', color='red')
ax2.tick_params('y', colors='red')

# Title and show the plot
plt.title('Average Boardings and Temperature by Hour of the Day')
plt.grid(False)
plt.tight_layout()
plt.show()
```
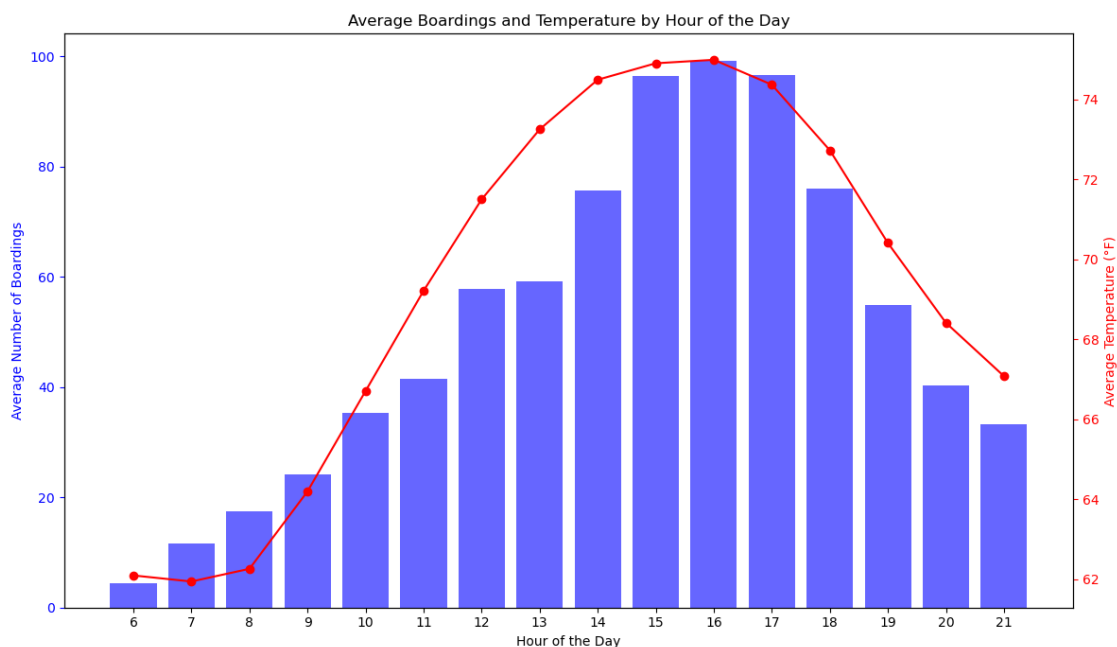
Insights:

The Average Boardings and Temperature by Hour of the Day uncovers an intriguing relationship. As the day progresses, the temperature tends to rise, peaking during midday. Concurrently, boardership also witnesses an uptick during these warmer hours. This mutual escalation gives an impression of a correlation between temperature and boardership. It's also possible that students and staff prefer to travel during the warmer parts of the day, or perhaps they're avoiding the chill of early mornings, especially during the fall semester. But given Austin's weather, that seems unlikely.

0.3   3. A Semester's Ebb and Flow

Average Boardings across Months :

```python
[10]: # Group by hour_of_day and month
grouped_hr_month = cm.groupby(['hour_of_day', 'month']).boarding.mean().
 ↪unstack().reset_index()

# Set up the figure
plt.figure(figsize=(12, 7))

# Plot for each month
months = ['Sep', 'Oct', 'Nov']
colors = ['blue', 'green', 'red']

for month, color in zip(months, colors):
    plt.plot(grouped_hr_month['hour_of_day'], grouped_hr_month[month],␣
 ↪color=color, marker='o', label=f'Average Boardings in {month}')

plt.xlabel('Hour of the Day')
plt.ylabel('Average Number of Boardings')
plt.title('Average Boardings by Hour for Each Month')
plt.xticks(grouped_hr_month['hour_of_day'])  # To ensure all hours are shown
plt.legend()
plt.grid(True, which='both', linestyle='--', linewidth=0.5)
plt.tight_layout()
plt.show()
```

Average Boardings by Hour for Each Month
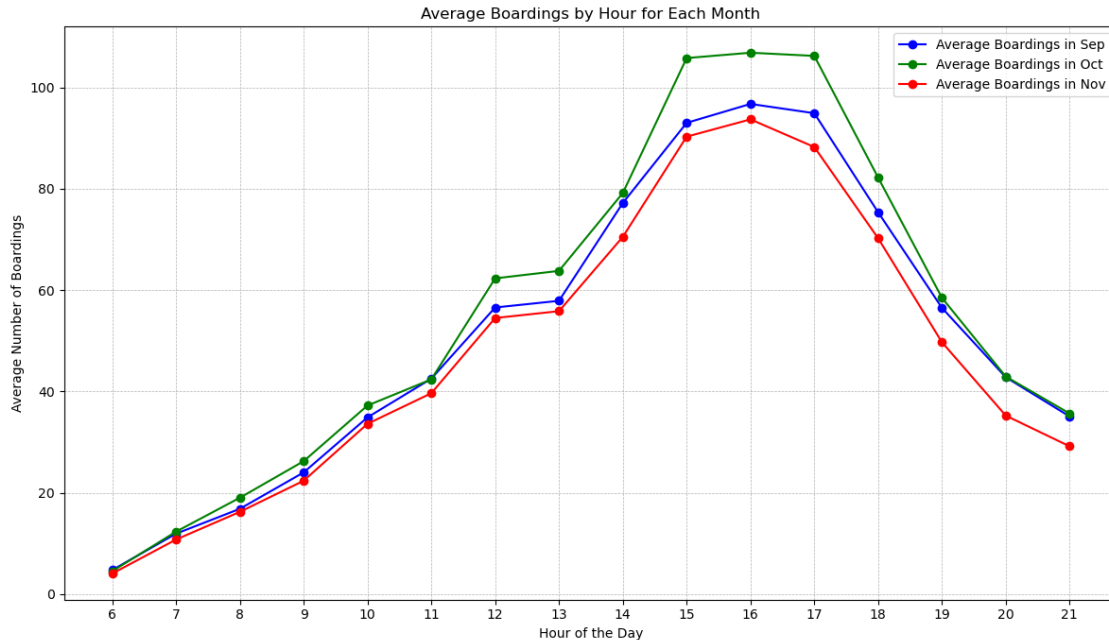
Insights: As we transition through months, the Average Boardings across Months reveals an arc in bus usage. Starting in September, the numbers rise in October, suggesting an increased reliance or familiarity with the bus system as the semester progresses. However, as November rolls around – perhaps as students bunker down for finals or projects – the numbers dip below even September levels. The distribution of boarding across hours of the day remains consistent across the months

0.4 4. Day-to-Day Fluctuations: Workdays versus Weekends

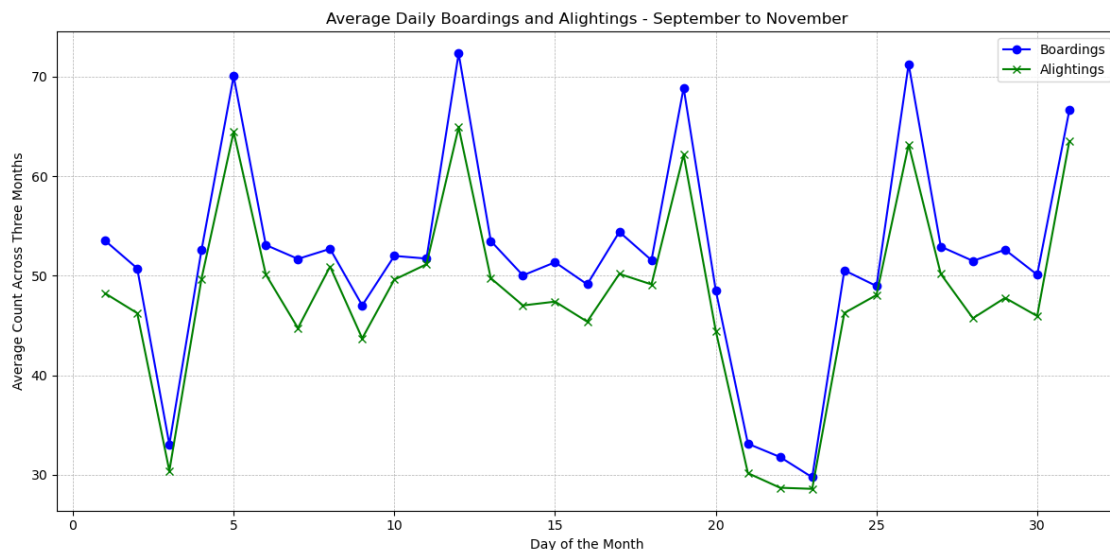Average Daily Boardings & Alightings : September to November

```python
[12]: # Extract day from timestamp
cm['day'] = pd.to_datetime(cm['timestamp']).dt.day

# Group data by day only
grouped_day = cm.groupby(['day']).agg({'boarding': 'mean', 'alighting':
 ↪'mean'}).reset_index()

# Plotting
plt.figure(figsize=(12, 6))

# Line plot for boardings and alightings
plt.plot(grouped_day['day'], grouped_day['boarding'], label='Boardings',
 ↪color='blue', marker='o')
plt.plot(grouped_day['day'], grouped_day['alighting'], label='Alightings',
 ↪color='green', marker='x')
```

```
plt.xlabel('Day of the Month')
plt.ylabel('Average Count Across Three Months')
plt.title('Average Daily Boardings and Alightings - September to November')
plt.legend()
plt.grid(True, which='both', linestyle='--', linewidth=0.5)
plt.tight_layout()
plt.show()
```


Insights:

The Daily Boardings & Alightings from September to November shows a cyclical pattern. There's a clear increase in boardings during weekdays and a decrease during weekends. Additionally, there's a noticeable reduction in boardings during weekends at the end of each month. This drop could be attributed to long weekends, holidays, or students taking breaks.

In conclusion, the Capital Metro data reveals distinct daily and monthly travel patterns at UT Austin during the fall semester, offering insights for transportation planning and campus activity.

[ ]: