Bike Rentals

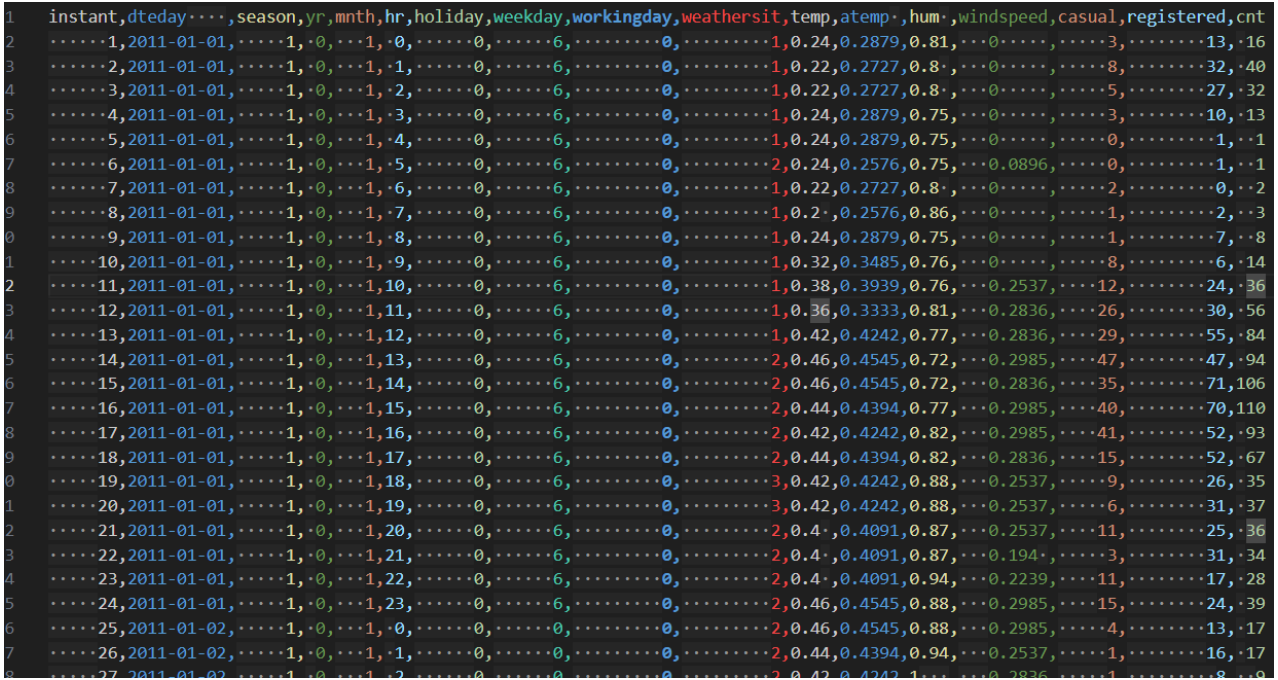
Data Analysis

Current Catalysts:

Vikram Anand, Umer Farooq

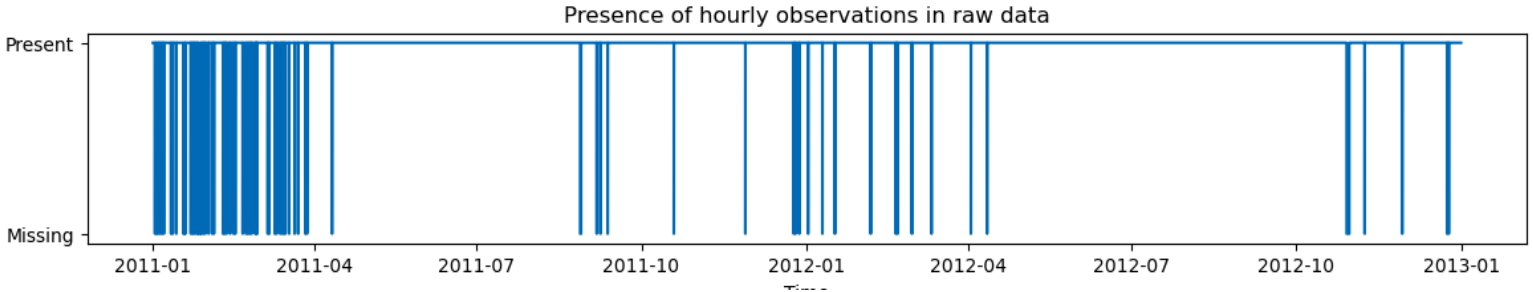
FH JOANNEUM
University of Applied Sciences

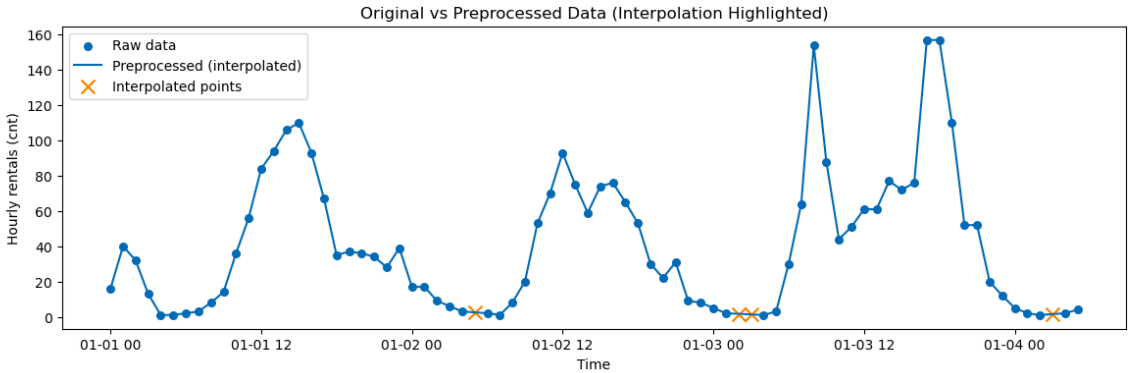
Data Set



```
instant,dteday,season,yr,mnth,hr,holiday,weekday,workingday,weathersit,temp,atemp,hum,windspeed,casual,registered,cnt
1,2011-01-01,1,0,1,0,0,6,0,1,0.24,0.2879,0.81,0,3,13,16
2,2011-01-01,1,0,1,1,0,6,0,1,0.22,0.2727,0.8,0,8,32,40
3,2011-01-01,1,0,1,2,0,6,0,1,0.22,0.2727,0.8,0,5,27,32
4,2011-01-01,1,0,1,3,0,6,0,1,0.24,0.2879,0.75,0,3,10,13
5,2011-01-01,1,0,1,4,0,6,0,1,0.24,0.2879,0.75,0,0,1,1
6,2011-01-01,1,0,1,5,0,6,0,2,0.24,0.2576,0.75,0.0896,0,1,1
7,2011-01-01,1,0,1,6,0,6,0,1,0.22,0.2727,0.8,0,2,0,2
8,2011-01-01,1,0,1,7,0,6,0,1,0.2,0.2576,0.86,0,1,2,3
9,2011-01-01,1,0,1,8,0,6,0,1,0.24,0.2879,0.75,0,1,7,8
10,2011-01-01,1,0,1,9,0,6,0,1,0.32,0.3485,0.76,0,8,6,14
11,2011-01-01,1,0,1,10,0,6,0,1,0.38,0.3939,0.76,0.2537,12,24,36
12,2011-01-01,1,0,1,11,0,6,0,1,0.36,0.3333,0.81,0.2836,26,30,56
13,2011-01-01,1,0,1,12,0,6,0,1,0.42,0.4242,0.77,0.2836,29,55,84
14,2011-01-01,1,0,1,13,0,6,0,2,0.46,0.4545,0.72,0.2985,47,47,94
15,2011-01-01,1,0,1,14,0,6,0,2,0.46,0.4545,0.72,0.2836,35,71,106
16,2011-01-01,1,0,1,15,0,6,0,2,0.44,0.4394,0.77,0.2985,40,70,110
17,2011-01-01,1,0,1,16,0,6,0,2,0.42,0.4242,0.82,0.2985,41,52,93
18,2011-01-01,1,0,1,17,0,6,0,2,0.44,0.4394,0.82,0.2836,15,52,67
19,2011-01-01,1,0,1,18,0,6,0,3,0.42,0.4242,0.88,0.2537,9,26,35
20,2011-01-01,1,0,1,19,0,6,0,3,0.42,0.4242,0.88,0.2537,6,31,37
21,2011-01-01,1,0,1,20,0,6,0,2,0.4,0.4091,0.87,0.2537,11,25,36
22,2011-01-01,1,0,1,21,0,6,0,2,0.4,0.4091,0.87,0.194,3,31,34
23,2011-01-01,1,0,1,22,0,6,0,2,0.4,0.4091,0.94,0.2239,11,17,28
24,2011-01-01,1,0,1,23,0,6,0,2,0.46,0.4545,0.88,0.2985,15,24,39
25,2011-01-02,1,0,1,0,0,0,0,2,0.46,0.4545,0.88,0.2985,4,13,17
26,2011-01-02,1,0,1,1,0,0,0,2,0.44,0.4394,0.94,0.2537,1,16,17
```

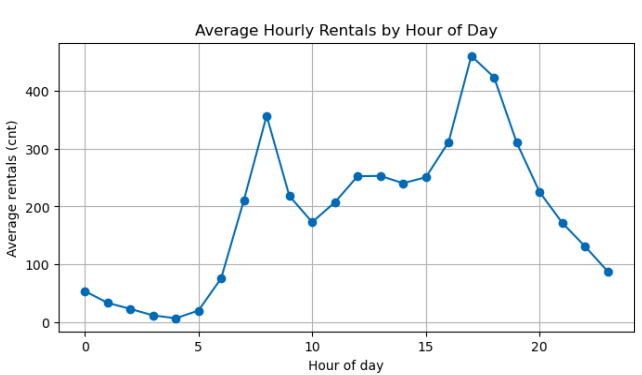
Missing Timestamps in Data and Interpolation

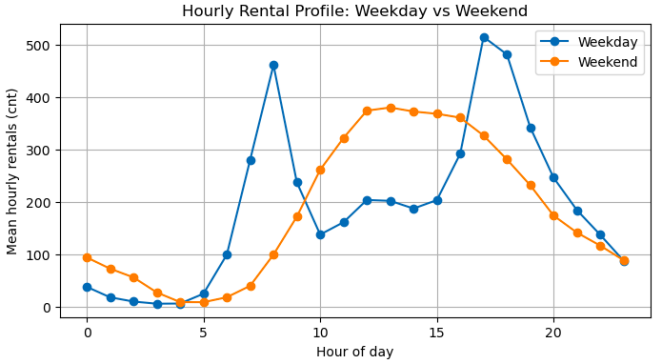


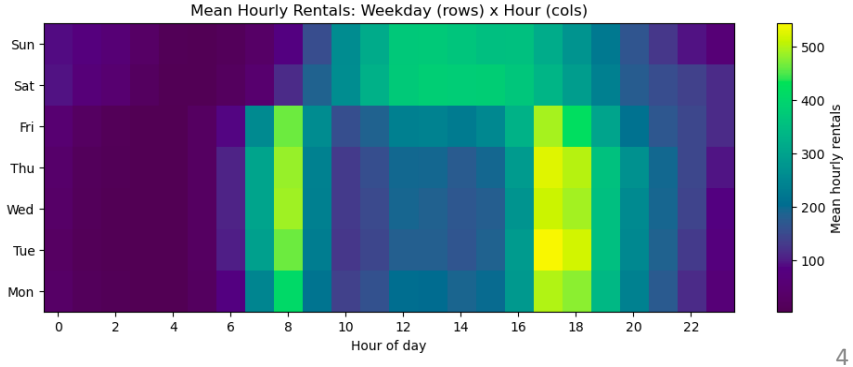


Mean Rentals of Bikes

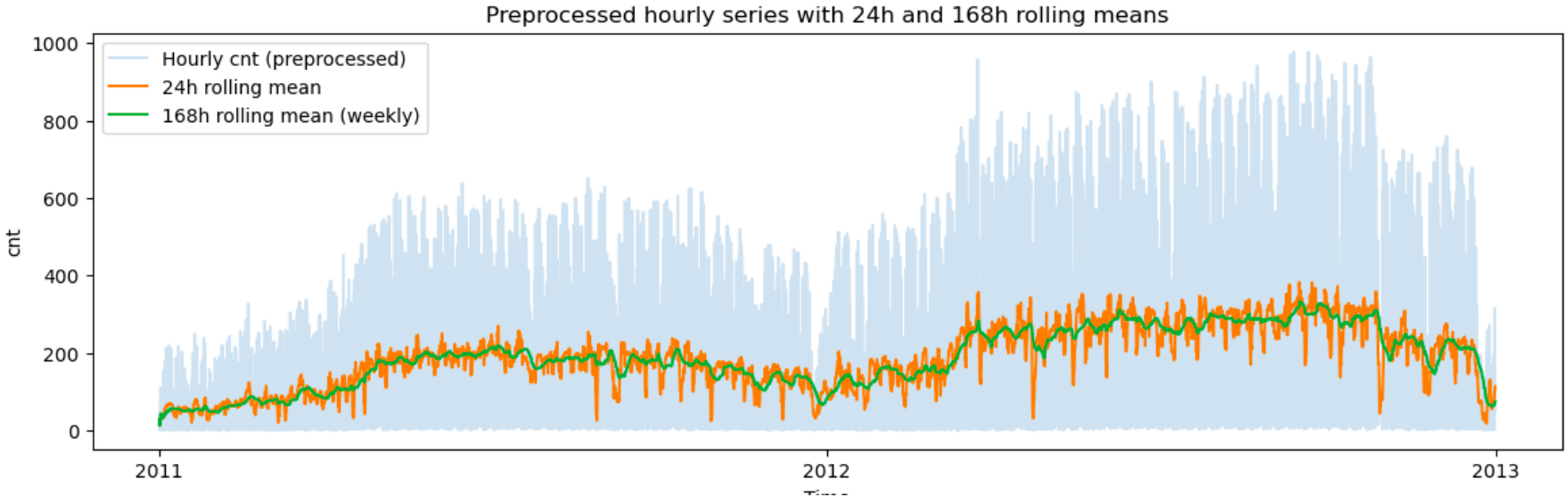
- Peak in rentals is observed during weekdays at
  - Morning hour around 0800.
  - During evening around 1800.
- Peak rentals is observed during weekend.
  - Between 1000 hrs and 1600 hrs.





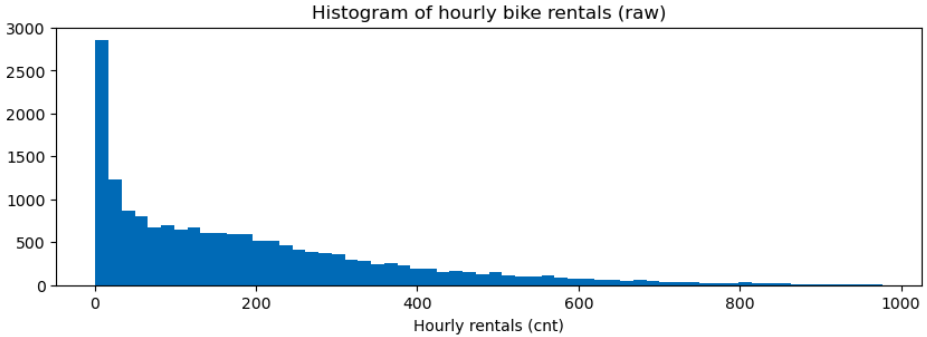


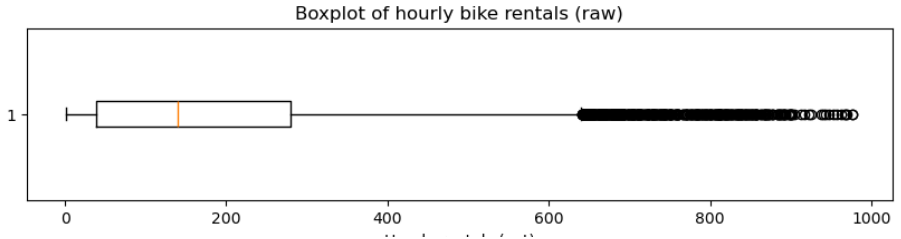
Preprocessing of data with 24h and 168h Rolling means



Histogram and Boxplot of Bike rentals.

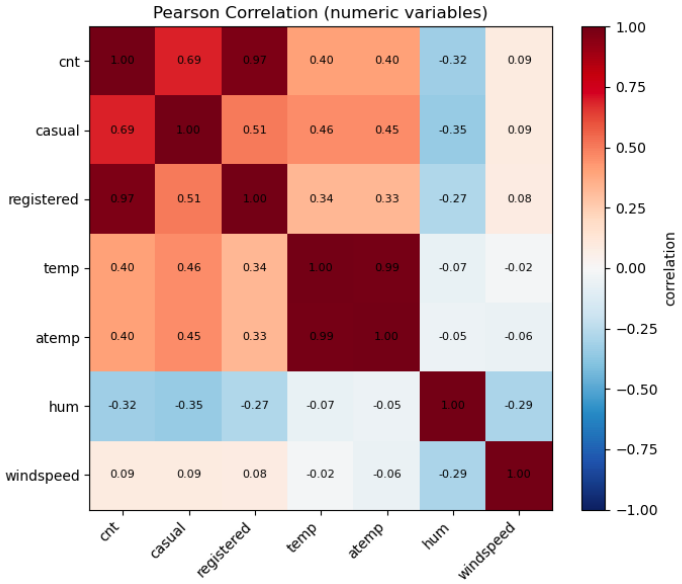
- Most of the data points are centered around at 0.
- Large number of rentals are observed very less time.
- Median of data is at around 140.
- A lot of Data points are above the maximum limit of outliers.
    - These data points are still covered in the data and are not filtered.
    - We are interested in the large numbers of rentals in particular hours.



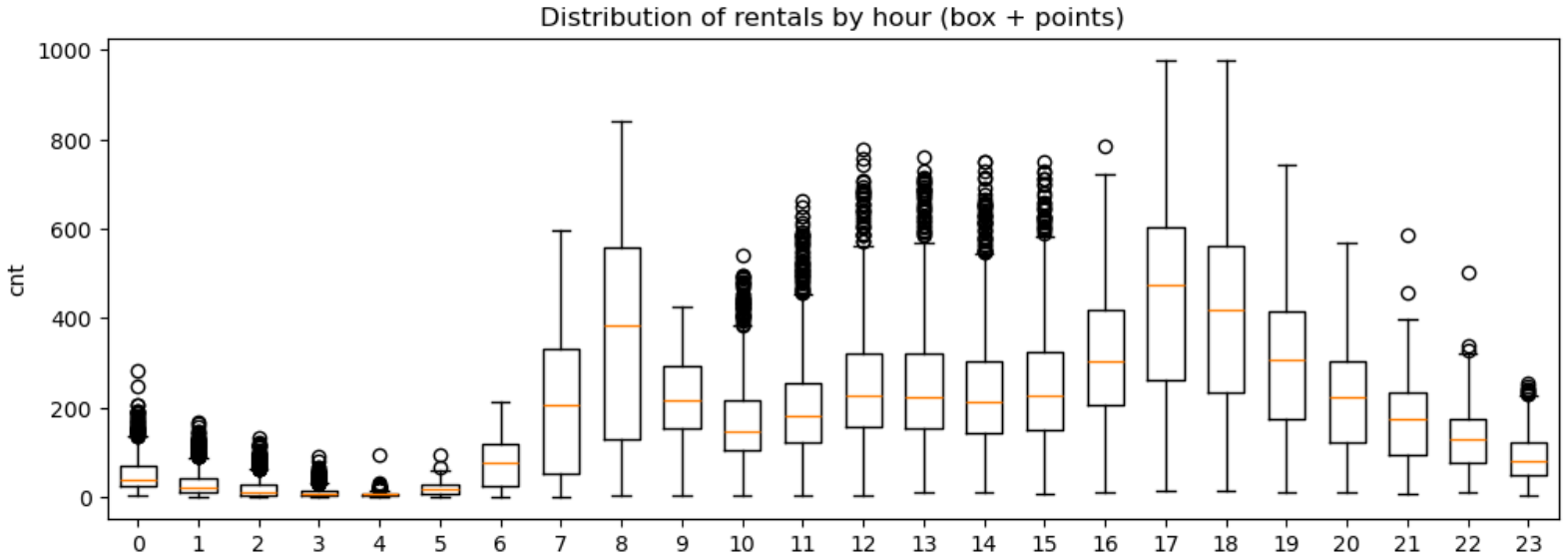


Pearson Correlation Matrix

- Pearson correlation matrix is used to observe the linear relationship between the number of hourly rentals and different parameters.
- Our only parameter of concern in temperature to analyze the data.
- We see that correlation appears to be 0.40 with the temperature which signifies a linear relationship between the hourly rentals and temperature.
    - If temperature increases (summers) the number of rentals are also increased.



Boxplot of rentals Per Day for Whole Data

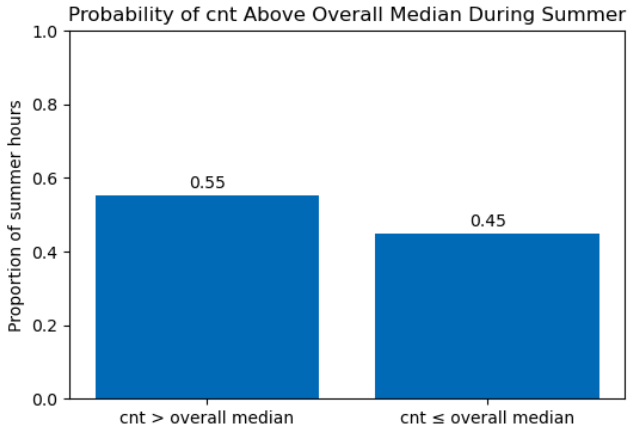


Probability of Rentals above overall median during Summers

- It is an Empirical probability
  - Because we are observing the real data and not the theoretical model.
  - Probability that rentals are above overall median given that it is summer.
  - Total number of hours in summer are 4409.
  - 2435 hrs the rentals are above overall median.
  - `P(cnt > overall median | summer) = 0.5523 (55.23%)`

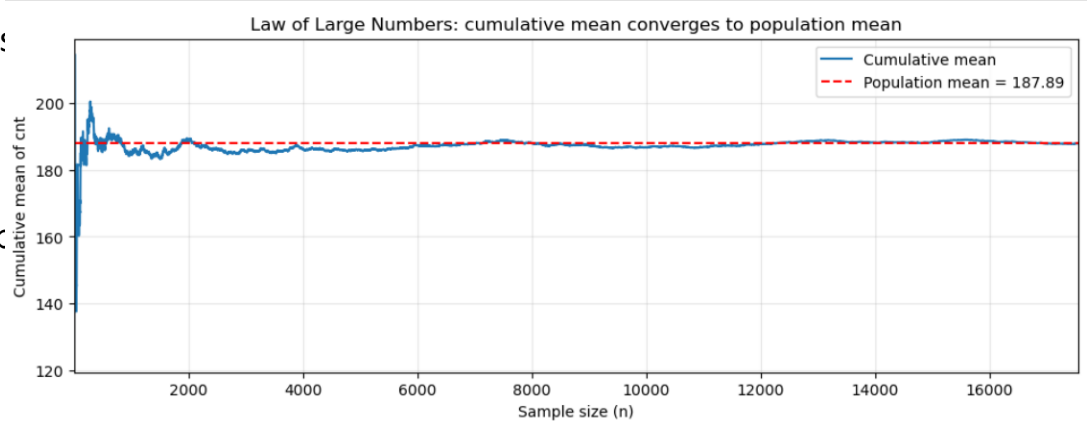
```
Overall median cnt = 140.0
Number of summer hours = 4409
Number of summer hours above overall median = 2435

Empirical probability:
P(cnt > overall median | summer) = 0.5523 (55.23%)
```



Law of Large Numbers

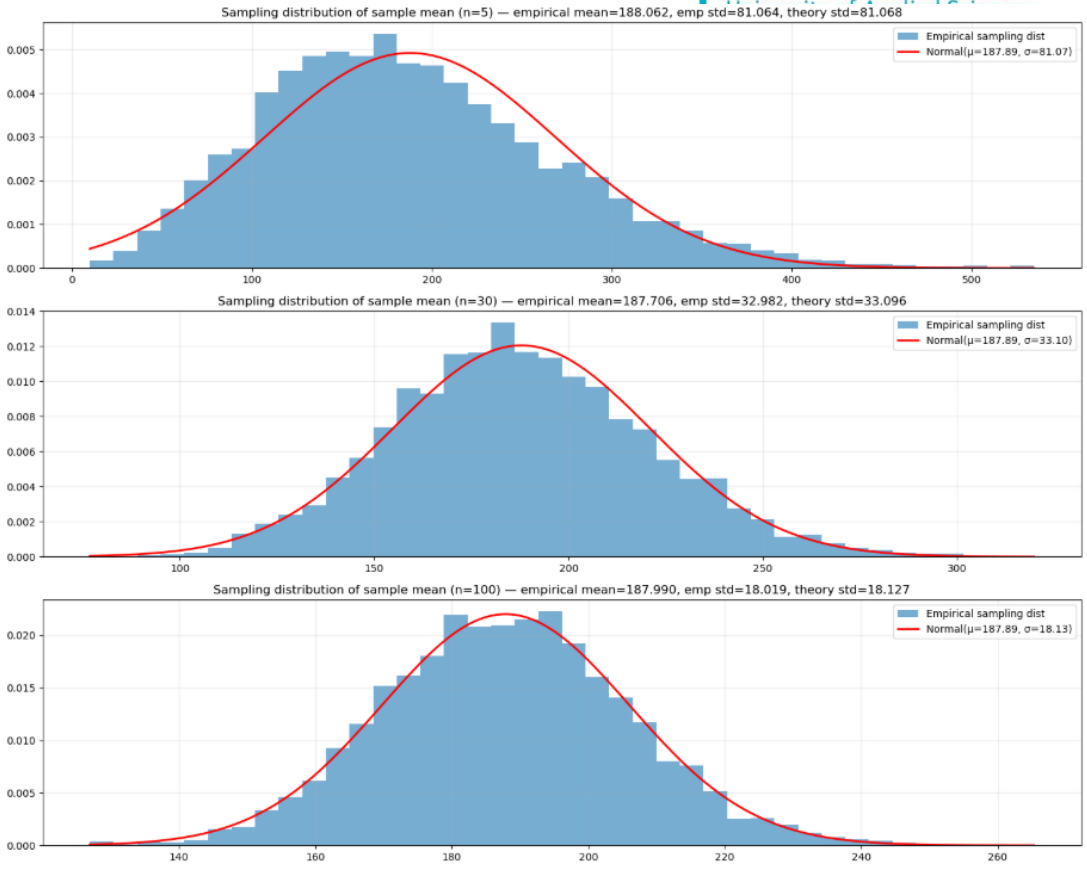
- As the number of observations increases, sample averages converge to the true population value.
- Cumulative probability of high-demand hours stabilizes as more hourly observations are included.
- Temporal dependence in hourly data slows convergence compared to independent samp but convergence is still observable.
- LLN justifies using long-term averages and empirical probabilities for stable demand estimation.



FH | JOANNEUM

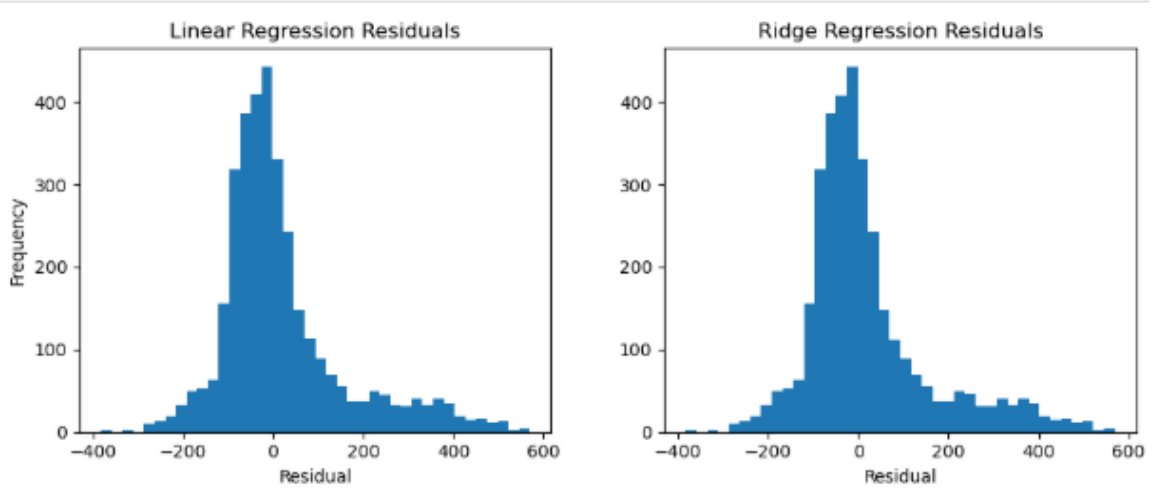
Central Limit Theorem

- The distribution of sample means approaches a normal distribution as sample size increases.
- sampling distributions of mean hourly rentals become more symmetric and concentrated with larger sample sizes.
- Larger samples reduce variability of the sample mean, improving stability and reliability of estimates.



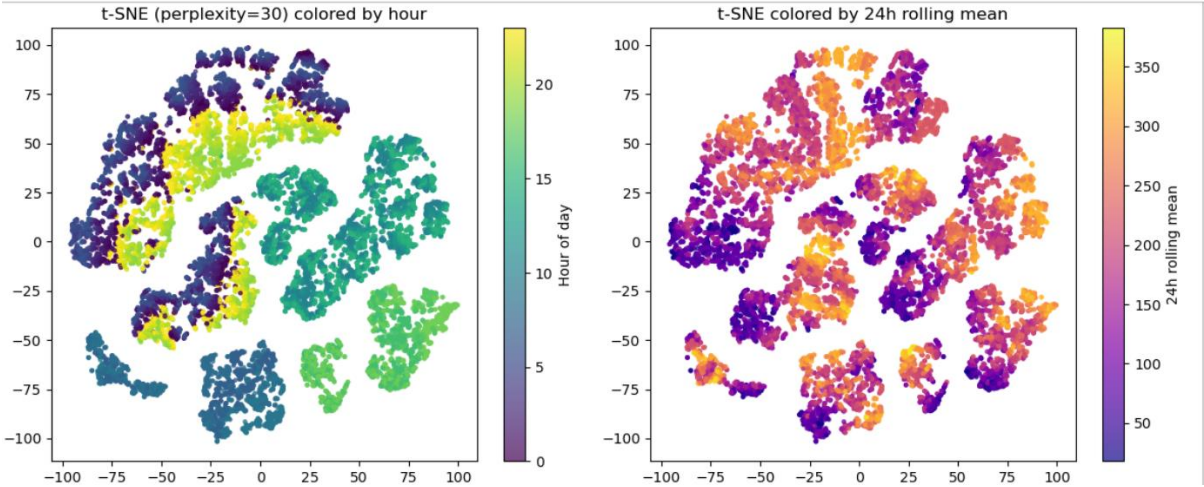
Residual Analysis

- Residuals are centered around zero, indicating no strong systematic bias.
- Most prediction errors are small, showing that the model performs well for typical demand level
- Most prediction errors are small, showing that the model performs well for typical demand level.



t-SNE

- The hour-of-day coloring transitions smoothly inside and across clusters, indicating that time of day is a major organizing factor.
- Each island groups hours with similar feature patterns.
- The rolling-mean coloring shows that clusters also separate by baseline demand intensity (low vs. high average usage)

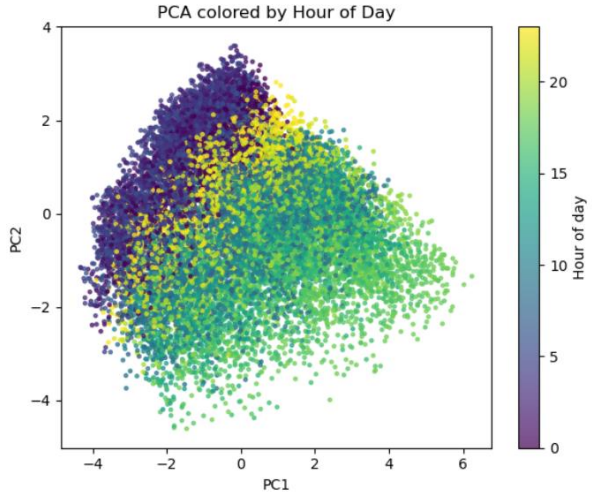


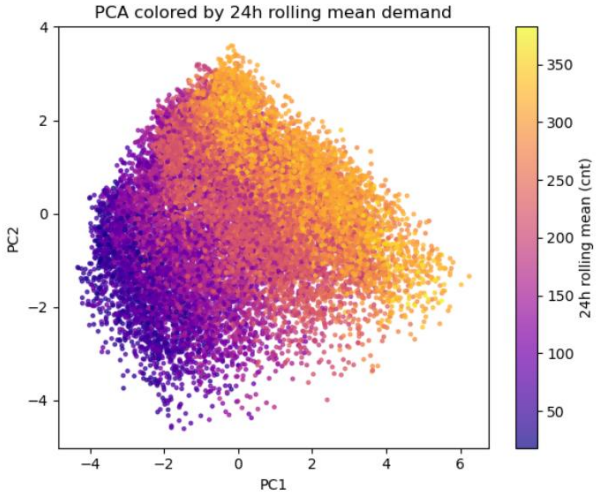
PC1 x PC2

PC plot shows the data projected onto the first two principal components, which capture the largest sources of variance in the dataset.

coloring by hour of day shows a smooth gradient, indicating that daily usage cycles are a major contributor to variance.

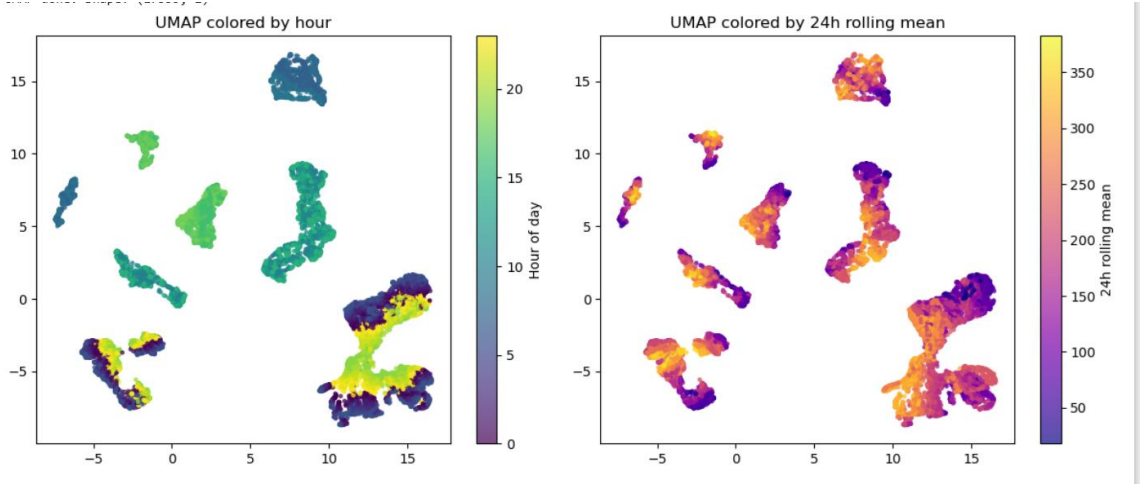
coloring by the 24-hour rolling mean demand shows an even clearer separation along PC1, suggesting that PC1 represents baseline demand intensity.





U-MAP

- UMAP produces clearer, more separated islands that preserve both local clusters and some global relationships
- UMAP produces distinct islands that correspond to groups of hours with similar features.
- Rolling-mean coloring clusters together regions of similar baseline demand — islands with warmer colors correspond to consistently higher-demand regimes



Q&A