Volvo Truck Analytics

Ioannis Batsios, William Downs, Wahab Ehsan, James Polk, and Christopher Thacker

General Overview

Basic Data Statistics and Analysis

- Utility functions (Wahab)
  - Display Data by Day
  - Estimator
  - Outlier Detection
- GPS Speed vs. Wheel-Based Speed (Chris)
- External Temperature vs. Components (Ioannis)
- CPU Load (James)
- APU Investigation (Bill)

Display Data per Day

Divided data by day using mean.

Function returns average for any attribute/column.

Truck 1 had one week worth data, but Truck 2 had only three days of data.

```
08/05/2019    77.935006
08/06/2019    73.576752
08/07/2019    76.170885
08/08/2019    74.289625
08/09/2019     3.091957
08/10/2019     2.655035
08/12/2019     2.966259
Name: Speed (km/hr), dtype: float64
```

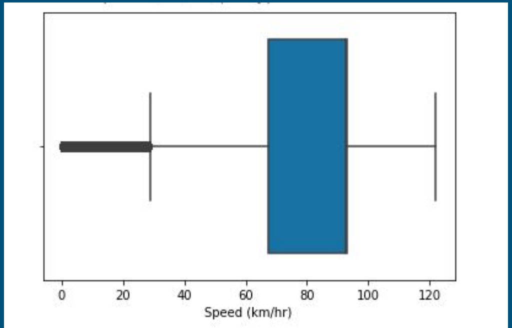
Truck 1

```
03/11/2019    25.192689
03/12/2019    19.167827
03/13/2019    25.641917
Name: Speed (km/hr), dtype: float64
```

Truck 2

Outlier Detection

Both trucks had several outliers making the data seem scattered.
Made function to show boxplot and then had it remove outliers.

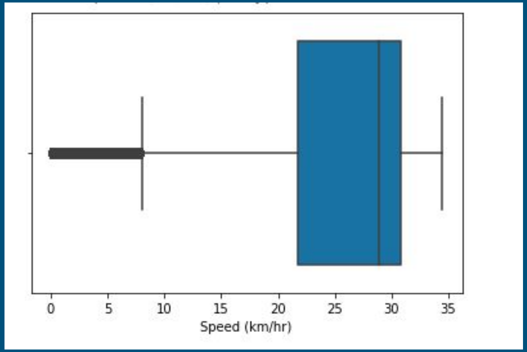


Truck 1

Truck 1

Before Outlier Deletion

Min:	0.0
Quartile 1:	67.22
Median:	92.60
Quartile 3:	93.15
Max:	122.05



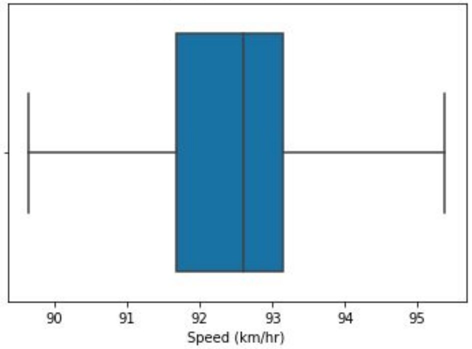
Truck 2

Truck 2

Before Outlier Deletion

Min:	0.0
Quartile 1:	21.71
Median:	28.91
Quartile 3:	31.87
Max:	34.47

```
After Outlier Deletion: [28.70599937438965, 91.67400360107422, 92.5999984741211, 93.15560150146484, 122.04680633544922]
Time (DateTime)
08/05/2019    88.711232
08/06/2019    86.577039
08/07/2019    86.438914
08/08/2019    84.607654
Name: Speed (km/hr), dtype: float64
```



Truck 1

<u>After Outlier Deletion</u>

Min: 28.71

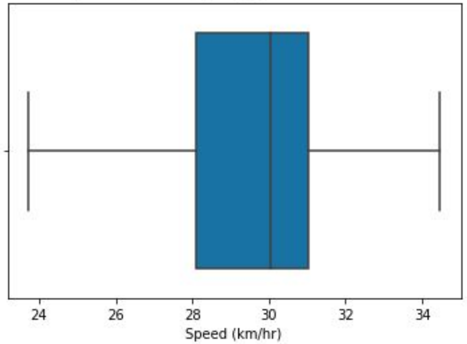
Quartile 1: 91.67

Median: 92.60

Quartile 3: 93.16

Max: 122.05

```
After Outlier Deletion: [8.025333616532958, 28.088665008769745, 30.043555365908862, 31.020999484840303, 34.467777676328375]
Time (DateTime)
03/11/2019    29.578152
03/12/2019    28.254743
03/13/2019    28.565267
Name: Speed (km/hr), dtype: float64
```



Truck 2

<u>After Outlier Deletion</u>

Min: 8.03

Quartile 1: 28.10

Median: 30.04

Quartile 3: 31.02

Max: 34.47

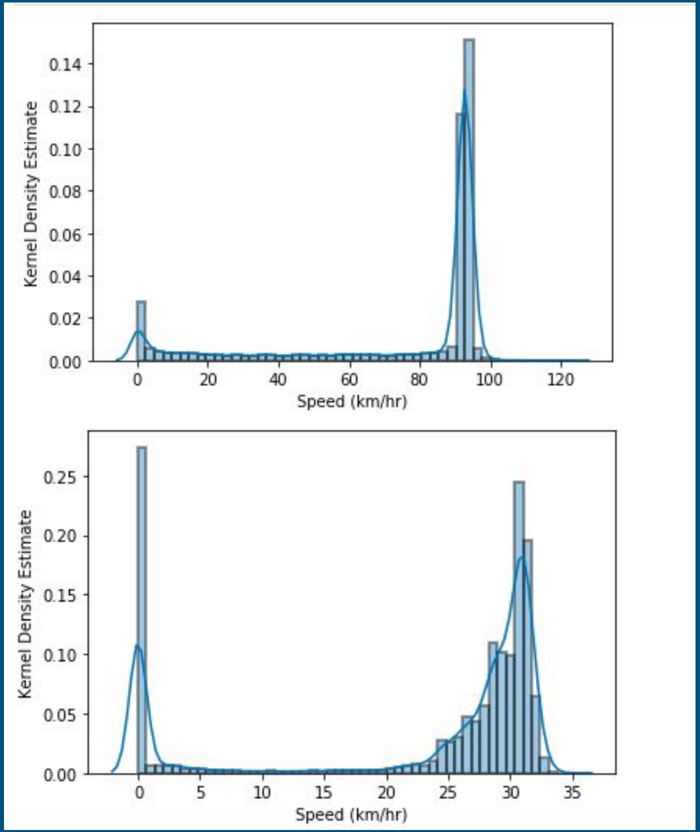
Determining the Estimator

Decided with Kernel Density Estimation (KDE).

Non-parametric estimator

There is no assumption for underlying distribution of variables.

Large Bandwidth since the data is mainly parsed around.



Truck 1

Truck 2

Oil Temperature

Ho: Temperature of the oil in both Trucks will remain the same.

Ha: Temperature of the oil in both Trucks will differ from each other.

Using Two Sample T-test with confidence interval of 95%.

```
Ttest_indResult(statistic=-30.33791875326352, pvalue=3.3950875287898653e-102)
```

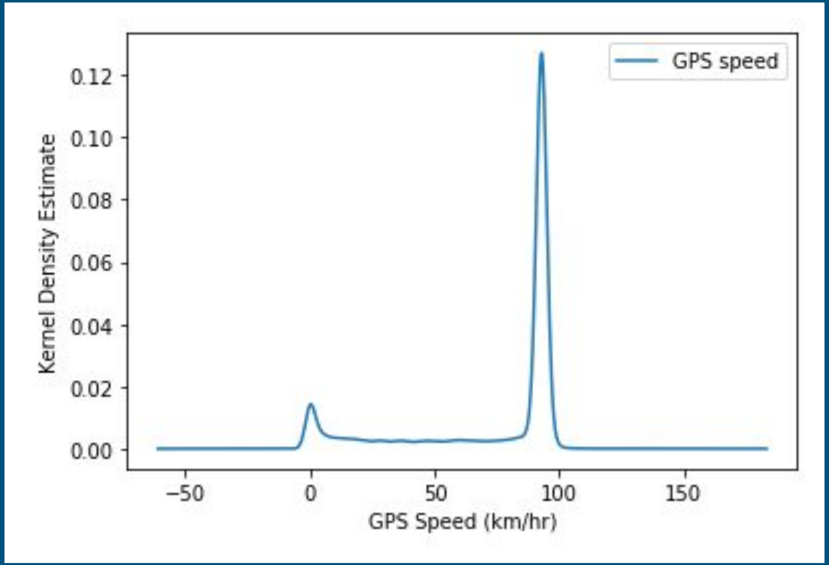
Reject the Null hypothesis. P-value less than 0.05. Therefore, there is difference in the Oil Temperature between trucks.

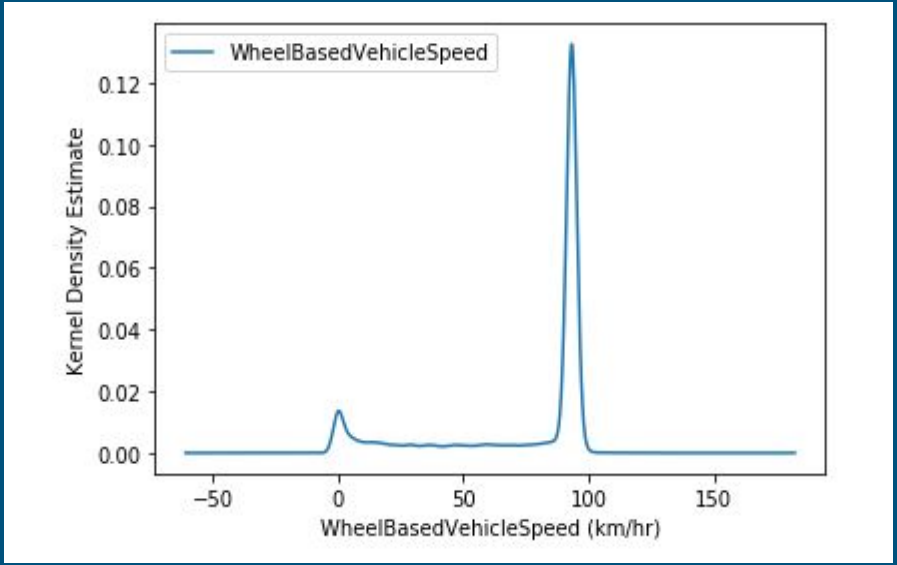
GPS Speed vs. Wheel-Based Speed

- Goals
  - Effectiveness of speed-measuring components.
  - Consistency between these components.
  - Basic understanding & exploration of data.
  - Basic data statistics.
  - Long-haul or short-haul?

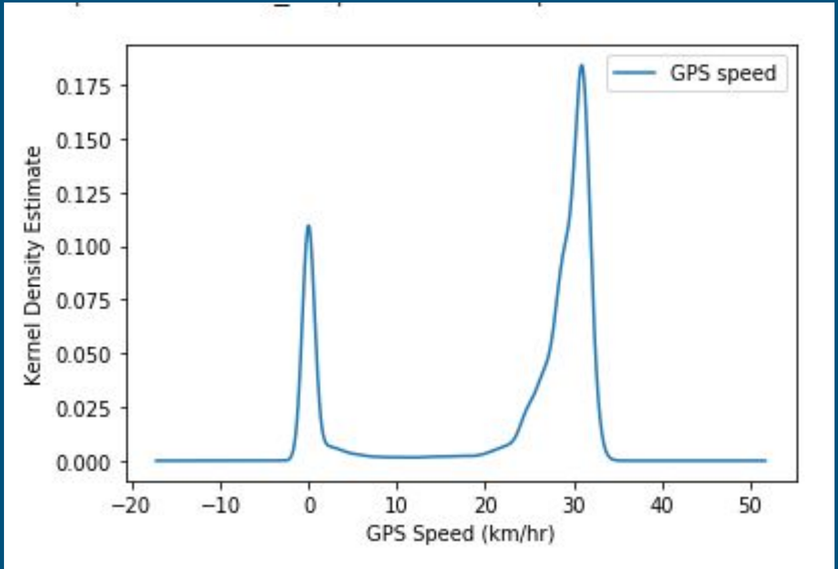
	Truck 1		Truck 2	
	GPS Speed	Wheel-Based Speed	GPS Speed	Wheel-Based Speed
Mean	74.55 km/hr (46.32 mph)	74.83 km/hr (46.50 mph)	22.69 km/hr (14.10 mph)	82.13 km/hr (51.04 mph)
Standard Deviation	31.94 km/hr (19.85 mph)	31.96 km/hr (19.86 mph)	12.18 km/hr (7.57 mph)	44.08 km/hr 27.39 mph

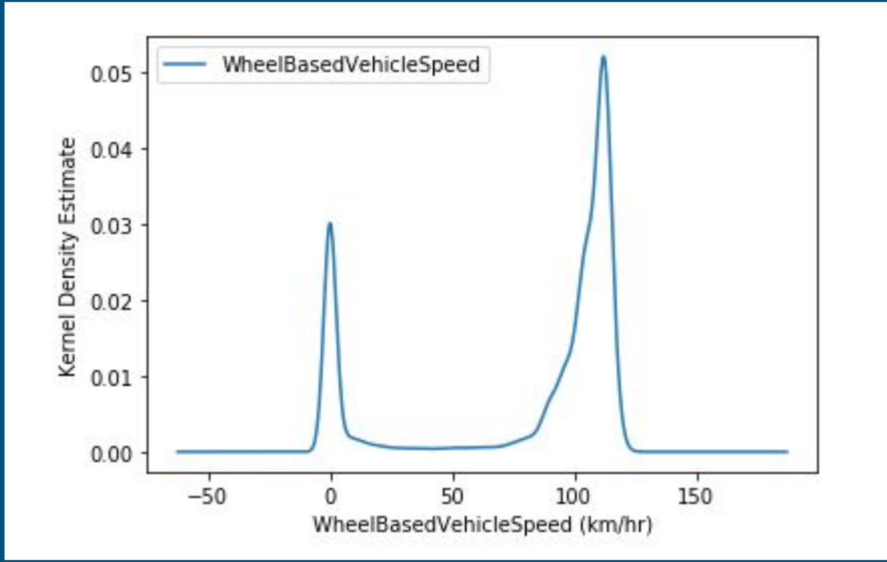
Truck 1: KDE Distributions





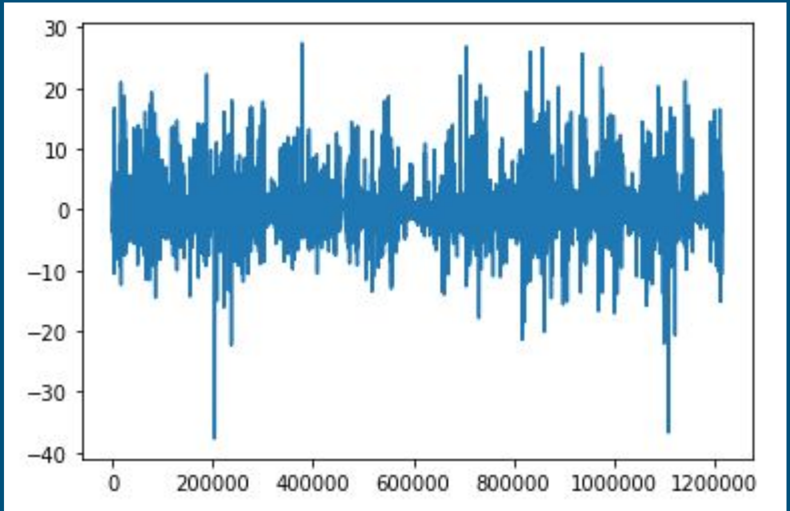
Truck 2: KDE Distributions



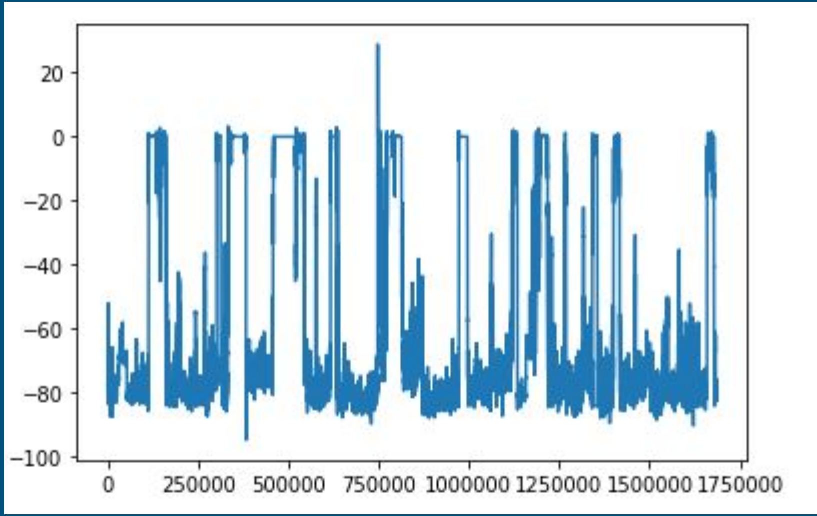


Difference in Speeds within Trucks

Truck 1



Truck 2



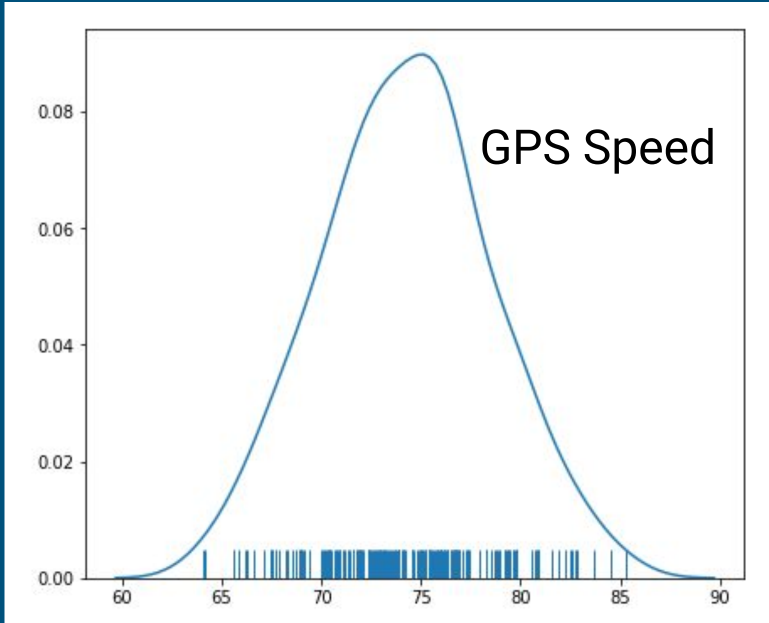
Hypothesis Testing for Speeds

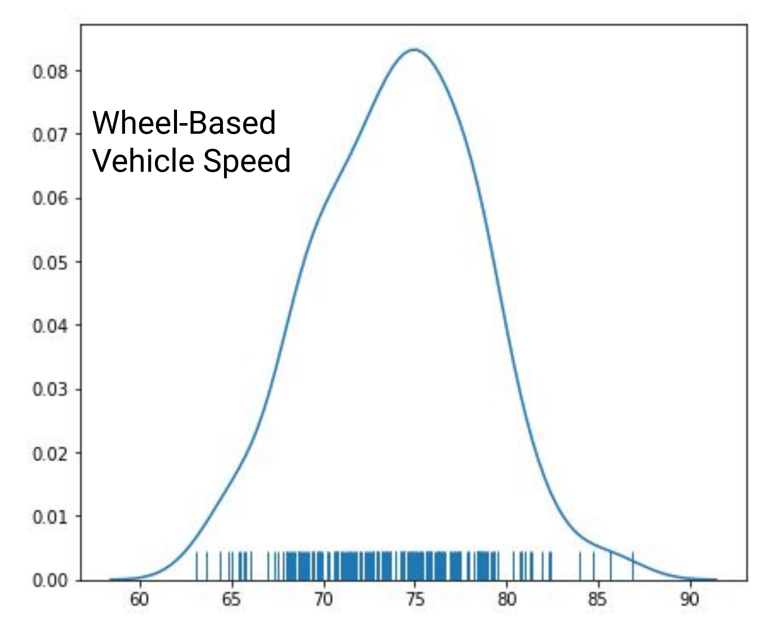
For these two speed components:

- A hypothesis test that will be performed is set up as:
  - H0: There are no differences in speeds between the two components.
  - Ha: There is a significant difference in speed between the two components.
    - Indicates two-tailed test (Scipy defaults to this).
- Since distribution is binomial:
  - Two-sample t-test will be used.
  - Central Limit Theorem required.
- **Though Truck 2 has faulty data, the same test will still be ran on Truck 2 for good measure but only Truck 1's test will be shown.**

Using the Central Limit Theorem for Truck 1

Truck 1: 200 Samples of Size 50 for Both Components





Two-Sample T-Test Results on Truck 1

Two-Sample T-Test: `Ttest_indResult(statistic=0.5068109121220089, pvalue=0.6125689000122868)`

- Confidence level of 95%.
  - Alpha level of 0.05.
- Using Scipy's built-in two-sample t-test:
  - P-value was ~0.613
  - Higher than alpha level of 0.05.

Thus, we fail to reject the null hypothesis and can assume that there is no significant difference between the measurements of both speed components for Truck 1.

Truck 2's P-value was lower than 0.05 (~1.12e-233) when the same test was ran. Thus, for Truck 2, we reject the null hypothesis and conclude that there is significant difference.

Temperature Analysis

I want to test my hypothesis that external temperature affects engine performance.

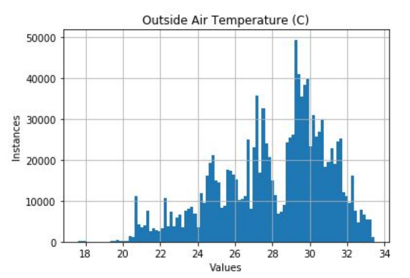
After some research, the main engine part that is affected by temperature is the turbocharger, which passes more air to the cylinder which can be mixed with more fuel to create more power. When the temperature is too hot, it is harder for the turbocharger to work resulting in less performance.

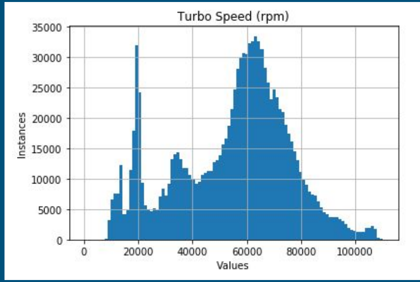
So the two columns I want to focus on are Turbo Speed and Outside Air Temperature

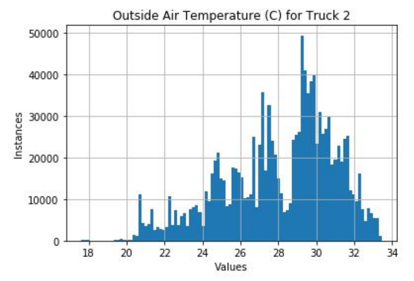
Null Hypothesis: Outside Air Temperature = Turbo Speed

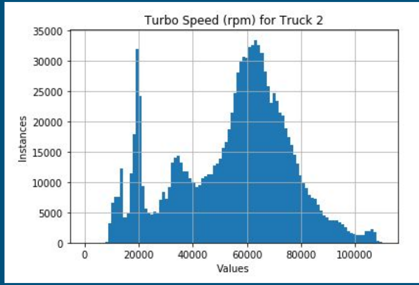
Alternative Hypothesis: Outside Air Temperature ≠ Turbo Speed

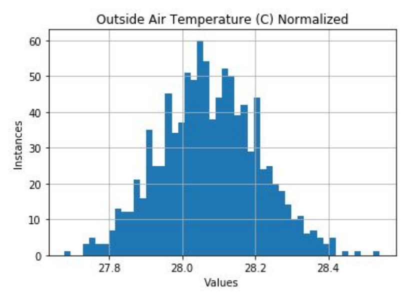
Temperature Analysis

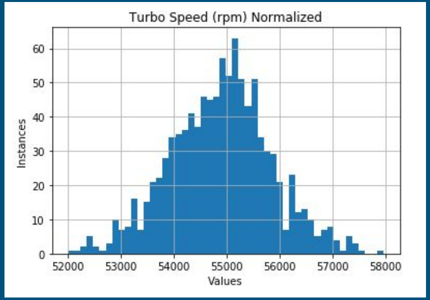


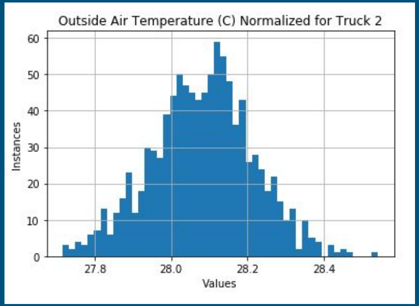


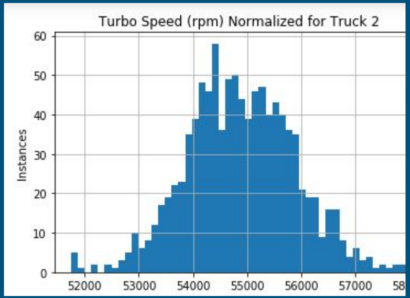








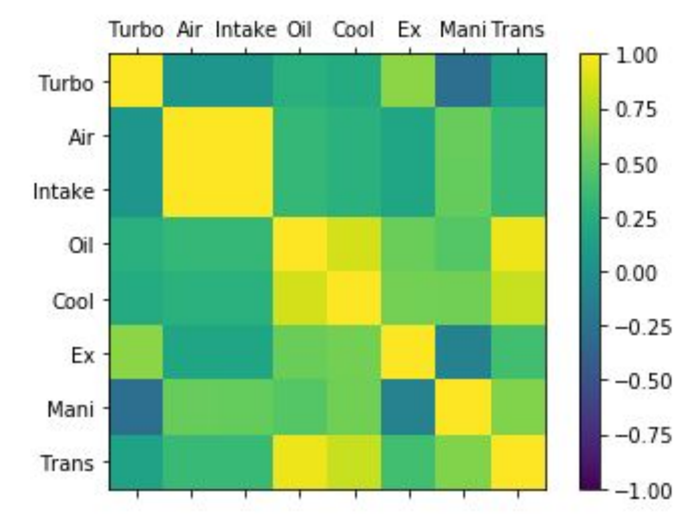




Temperature Analysis

	tse_TurboSpeed	AmbientAirTemperature_V
tse_TurboSpeed	1.000000	0.051154
AmbientAirTemperature_V	0.051154	1.000000

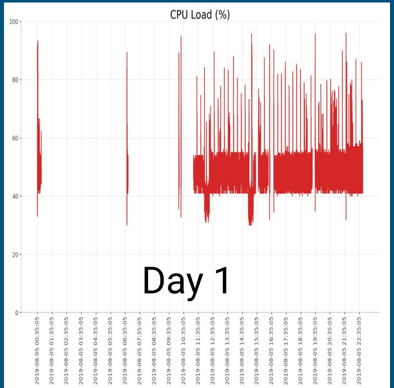
```
In [29]: stats.pearsonr(turbo_points, air_points)
Out[29]: (0.03713754670861208, 0.24066419089054117)
```

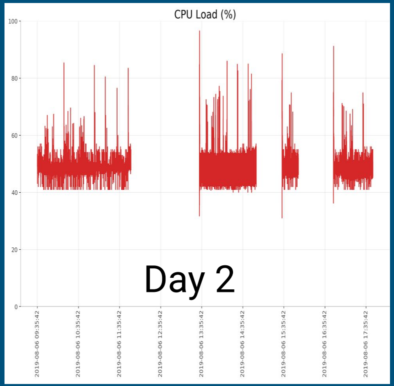


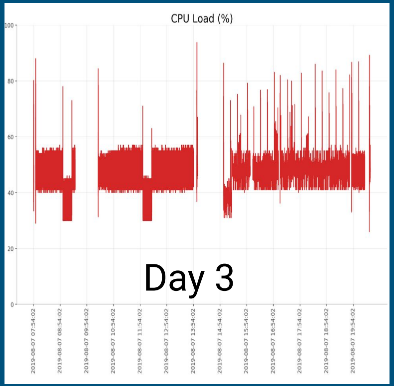

```
stats.pearsonr(turbo_points2, air_points2)
(0.019906520945919253, 0.5294967351973996)
```

The p-value is greater than .05, therefore, so we accept the null hypothesis that the outside air temperature does affect the turbo speed.

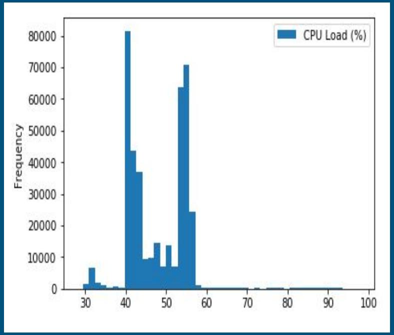
CPU Load

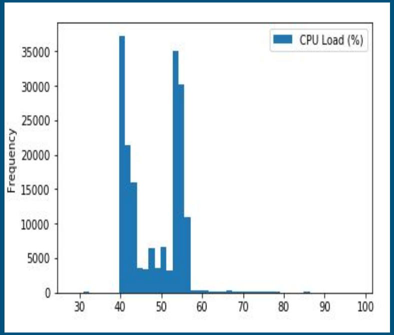


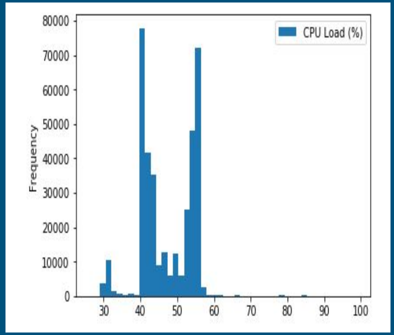


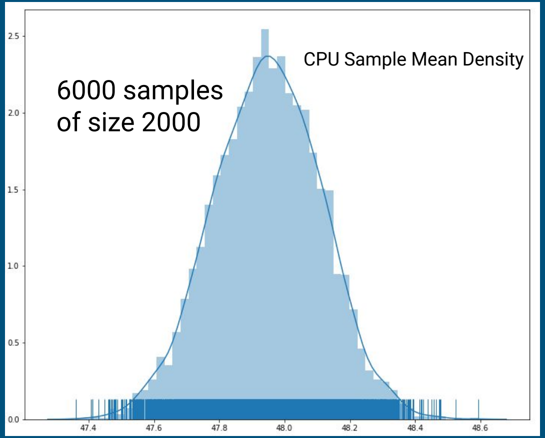


CPU Load (%)	
count	1,200,861
mean	47.94898
std	7.225541
min	26.00000
25%	42.00000
50%	47.00000
75%	54.00000
max	99.00000

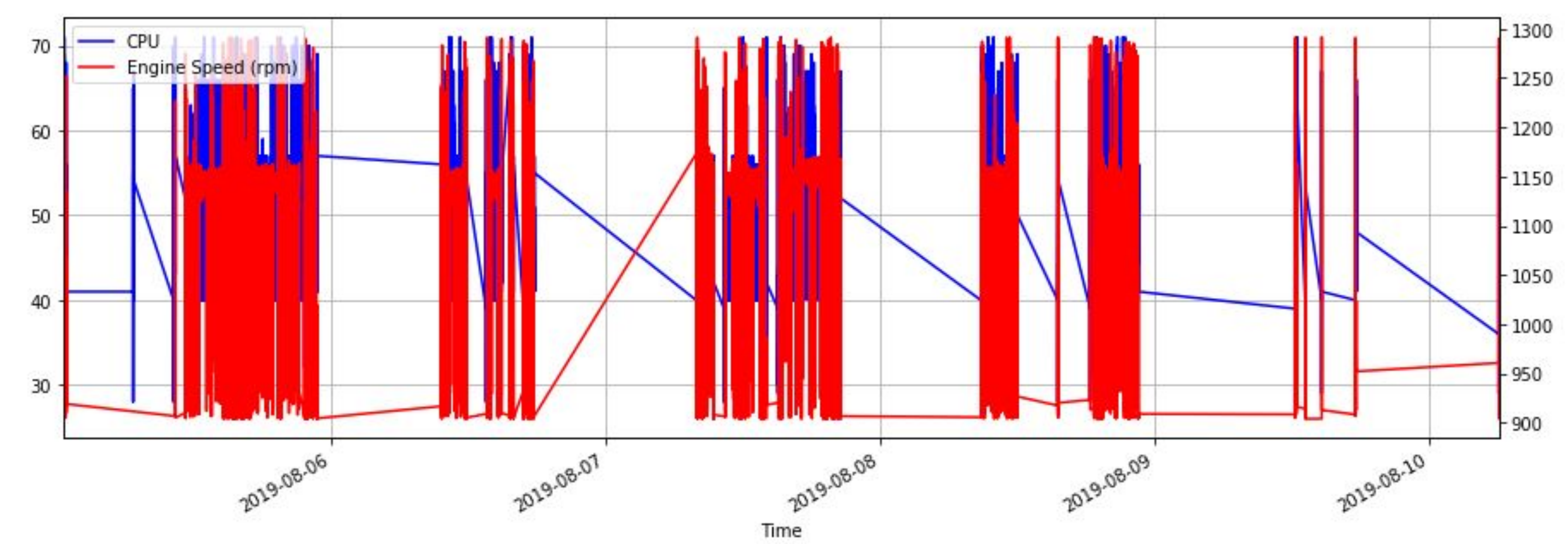






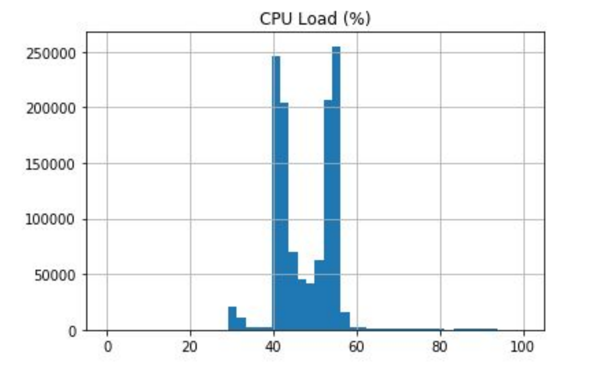


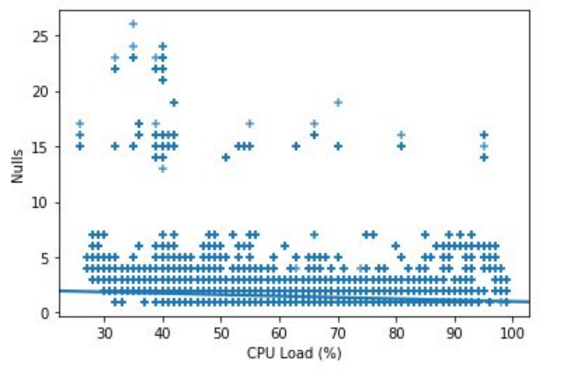
CPU Load: Why the Gaps?



CPU Load - Linear Regression

Is the CPU Load related to the number of active sensor readings?





Correlation:
-0.11680395911398092

R2 values using k-folds cross-validation (k=3):
0.01438574935982073,
0.013693048287461984,
0.012799950799256221

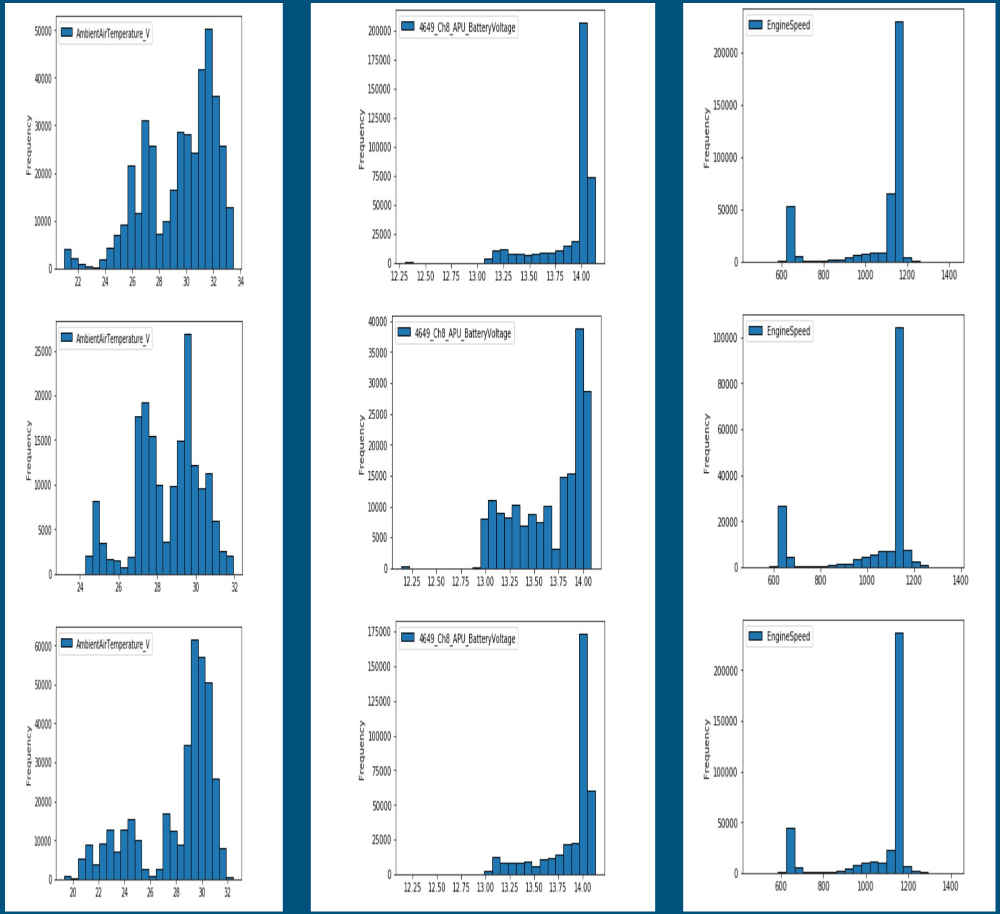
No significant correlation!

Single day distributions APU voltage and Engine speed

Auxiliary Power Unit

APU is a unit that is utilized to run comfort systems on the truck while the driver is on break, minimizing idle time fuel consumption. Traditionally all electrical power is derived from the alternator while idling at truck stops. Does the APU help any? Gaining insights on sensors associated with the apu may help determine if parasitic losses are worth having an APU unit.

12 Volt system supplied by 3 deep cells in parallel.



3 day and single day means, standard deviations

3 day describe

	4649_Ch8_APU_BatteryVoltage	AmbientAirTemperature_V	\
count	950905.000000	951142.000000	
mean	13.842501	28.768137	
std	0.303386	2.699031	
min	12.140536	19.312500	
25%	13.738613	27.218750	
50%	14.006409	29.468750	
75%	14.041657	30.625000	
max	14.130007	33.468750	

	4649_Ch1_Alternator_250A	EngineSpeed
count	950905.000000	951150.000000
mean	41.727043	1052.641237
std	30.472934	179.300508
min	3.304532	461.500000
25%	22.874228	1060.375000
50%	27.375639	1142.500000
75%	49.653811	1146.875000
max	232.475967	1420.500000

August 5th

	4649_Ch8_APU_BatteryVoltage	AmbientAirTemperature_V	\
count	401504.000000	401605.000000	
mean	13.903975	29.521066	
std	0.264101	2.616100	
min	12.305333	20.968750	
25%	13.897002	27.437500	
50%	14.032502	30.187500	
75%	14.047150	31.593750	
max	14.130007	33.468750	

	4649_Ch1_Alternator_250A	EngineSpeed
count	401504.000000	401610.000000
mean	36.722244	1054.688537
std	28.268899	178.054019
min	5.087930	461.500000
25%	22.569047	1097.625000
50%	25.630388	1142.250000
75%	33.288510	1145.375000
max	226.601244	1420.500000

August 6th

	4649_Ch8_APU_BatteryVoltage	AmbientAirTemperature_V	\
count	180841.000000	180876.000000	
mean	13.653022	28.527477	
std	0.352532	1.687976	
min	12.140536	24.437500	
25%	13.345388	27.281250	
50%	13.792630	28.875000	
75%	13.959258	29.687500	
max	14.081025	31.937500	

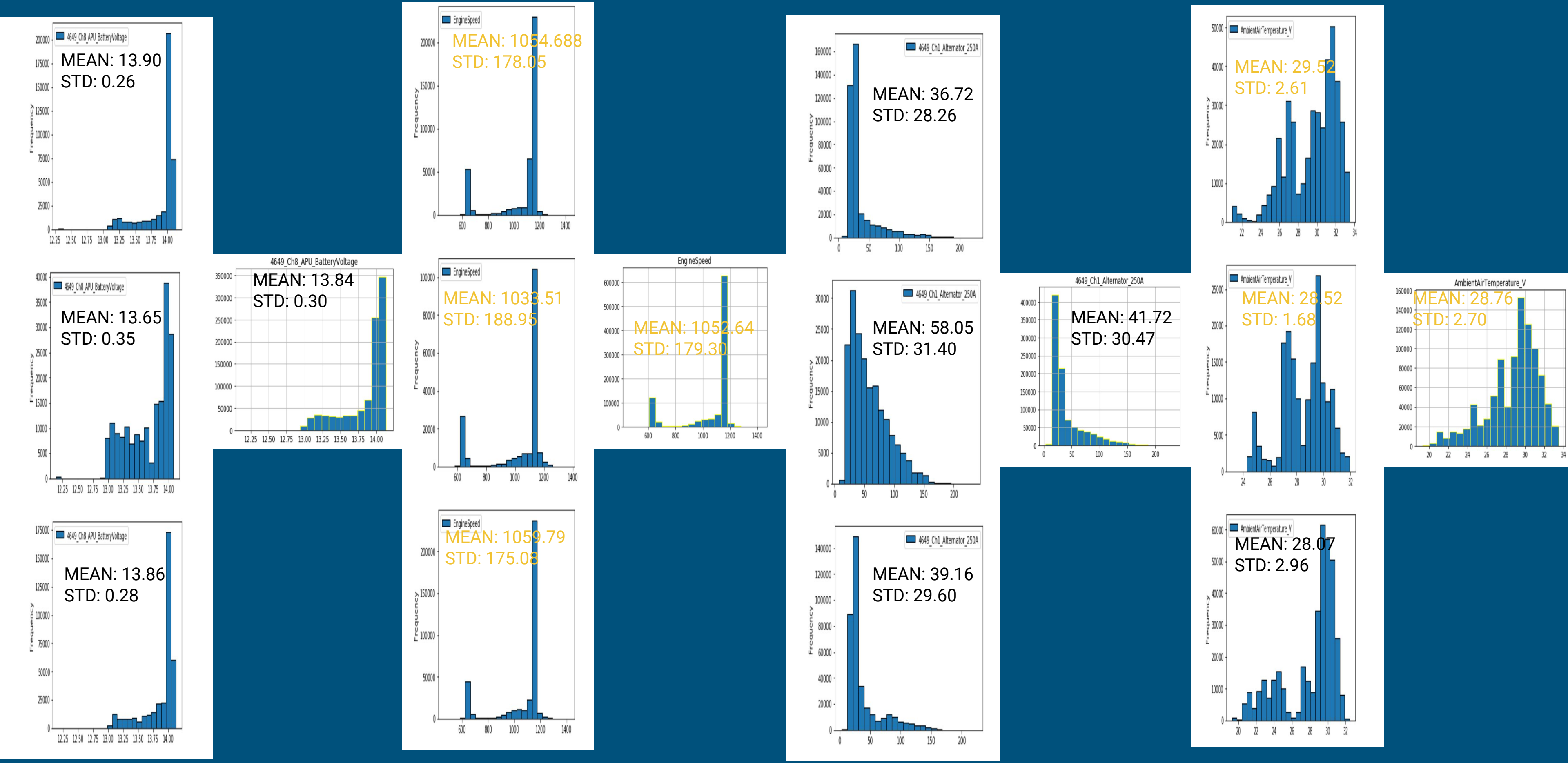
	4649_Ch1_Alternator_250A	EngineSpeed
count	180841.000000	180877.000000
mean	58.055357	1033.517121
std	31.405827	188.950453
min	7.262341	510.500000
25%	32.115473	1007.000000
50%	49.844549	1141.000000
75%	76.614595	1145.000000
max	232.475967	1368.375000

August 7th

	4649_Ch8_APU_BatteryVoltage	AmbientAirTemperature_V	\
count	368560.000000	368661.000000	
mean	13.868505	28.066001	
std	0.279471	2.966850	
min	12.172580	19.312500	
25%	13.805447	26.906250	
50%	14.011444	29.312500	
75%	14.040742	30.156250	
max	14.129091	32.437500	

	4649_Ch1_Alternator_250A	EngineSpeed
count	368560.000000	368663.000000
mean	39.167400	1059.793829
std	29.608507	175.086924
min	3.304532	466.750000
25%	21.739338	1064.625000
50%	26.259823	1143.875000
75%	41.576066	1154.750000
max	223.320554	1412.500000

3 day distributions
Vs single day

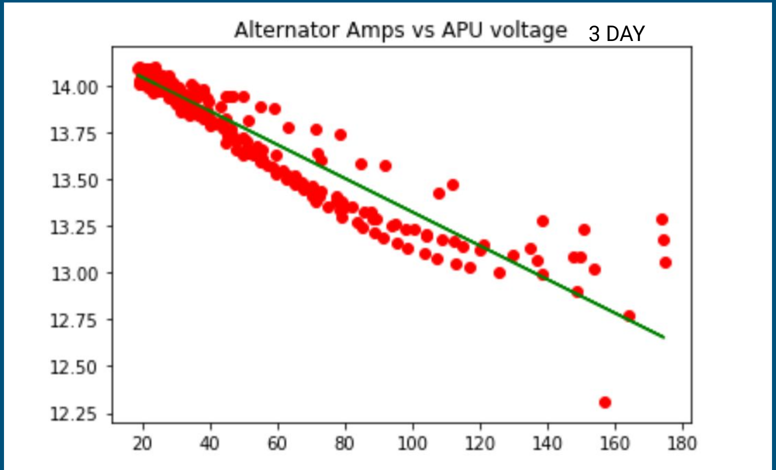


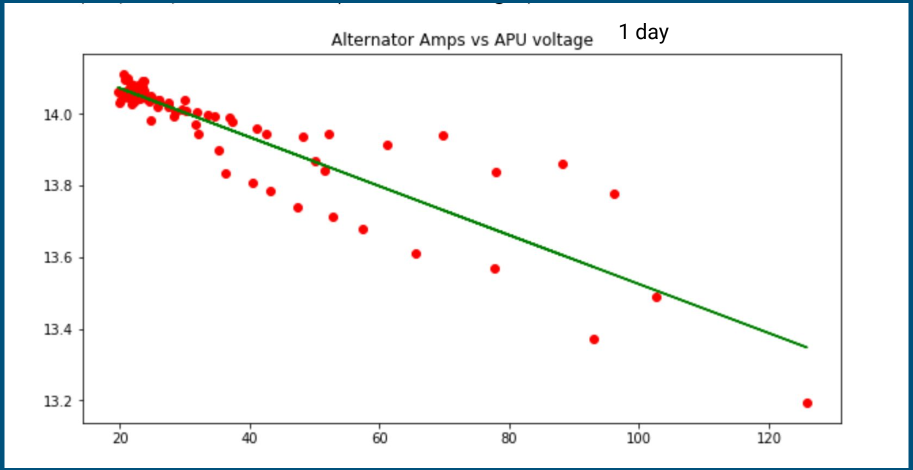
APU: Investigations and ideas!

At any given time when the engine speed is less than idle(off) the apu batteries should be discharging.

Any time the engine speed is greater than idle the batteries should be charging or charged.

As batteries discharge they typically will have a drop in voltage. Lets see a scatter plot of alternator vs battery voltage





Correlation

APU/Alternator correlation matrix

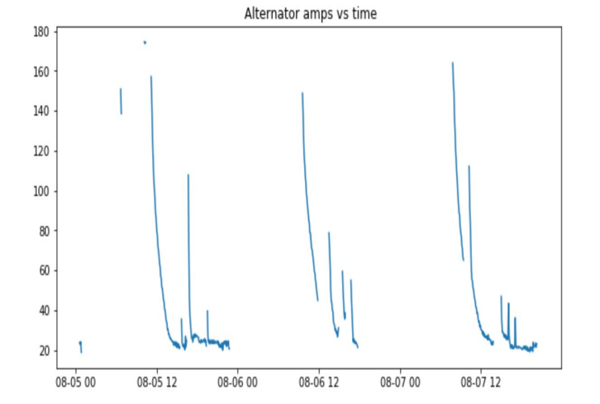
```
[[ 1.         -0.94908348]
 [-0.94908348  1.        ]]
```

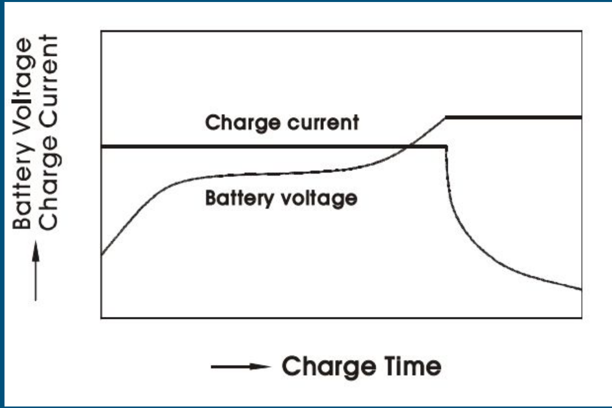
APU/Alternator Covariance matrix

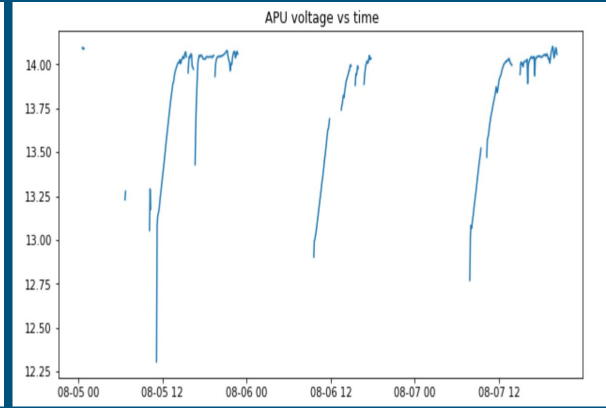
```
[[ 1.03068101e-01 -1.03081133e+01]
 [-1.03081133e+01  1.14452494e+03]]
```

This shows a strong negative relationship for the APU unit and the alternator (charging)

Lead Acid battery charging cycle with Time Series



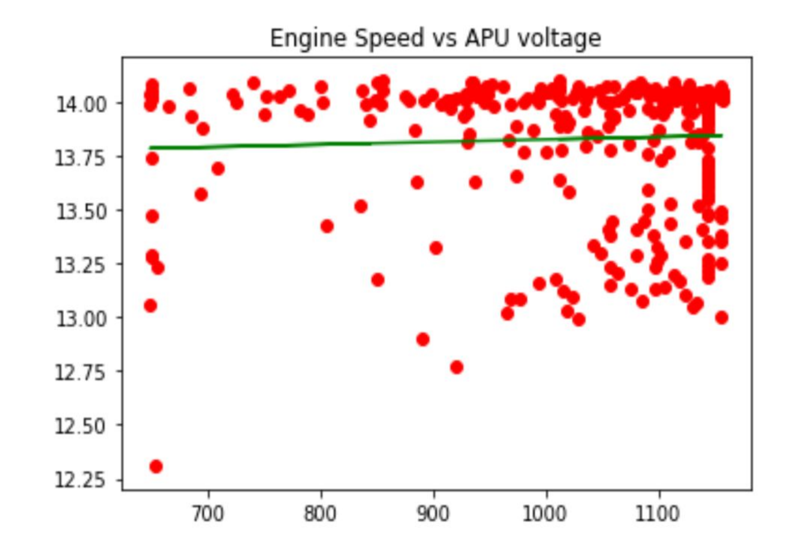


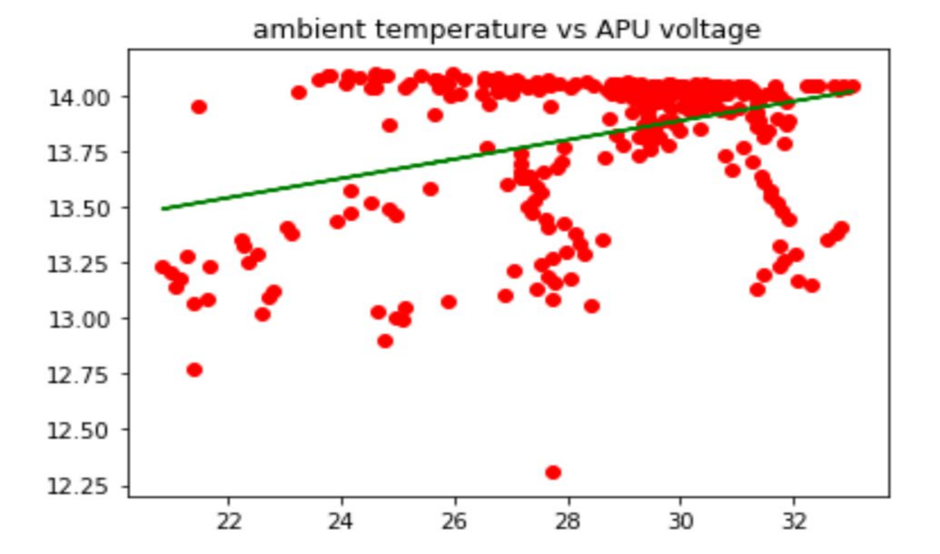


Source: DigiKey

https://www.digikey.com/eewiki/display/Motley/Charging+Lead+Acid+Battery+Basics

More scatter plots w fitted line





Since voltage and amperage are related, investigating temperature and engine speed vs APU to see if there is any correlation. The lack of correlation between these two metrics and the apu will need more discussion with Volvo, as we would assume we would see correlation with engine speeds and APU and maybe slight correlation with temperature.

*APU may be temperature controlled *Engine speed may be a different metric than what is assumed by the name.

Ambient Temp and Engine Speed vs APU Correlation

```
******Ambient temp and APU voltage******
correlation
 [[1.         0.36553273]
 [0.36553273 1.        ]]

covariance
 [[0.1030681  0.31816982]
 [0.31816982 7.35091062]]

******Engine speed and APU voltage******
correlation
 [[1.         0.05023186]
 [0.05023186 1.        ]]

covariance
 [[1.03068101e-01 2.20561877e+00]
 [2.20561877e+00 1.87058778e+04]]
```

1 Tailed T testing and P-value

Null hypothesis: The 3 day sample mean from the APU voltage is 13.82 is higher than the population APU Voltage mean.

The calculated p-value is -6.11e-13 which is well less than .05 significance level, so we can reject the null hypothesis indicating that with 95% confidence the sample mean will be lower than the population mean.

Errors and more

APU: Need a metric that only operates when alternator is off. Investigation into data, as wells as discussions with Volvo engineers revealed that the data may not include sensors that accurately isolate the APU.

APU: A capacity metric would be nice.

GPS and Wheelbase speed outliers.

Other outliers when truck is stopped or when a sensor first starts.