Predicting Lateness of SEPTA Regional Rail

THE PROBLEM?

SEPTA estimates a train's arrival time via its TrainView application using a simple procedure that produces inaccurate predictions of lateness

(Estimated Arrival) = (Status)+(Scheduled Arrival)

What's missing:

1. External influences
2. Predicting Early Arrivals
3. Estimating arrival times of future trips

OUR SOLUTION

- Take into account external influences
- Test Gaussian Process Regression, Linear Regression, and Random Forest models on SEPTA data
- Allow passengers to predict lateness for any destination along regional rails at any time

Train ID	Distance from Origin	Pause at Station	Time of Day	Day of Week	Month
1	2 (stations)	33 (seconds)	43,200	7	4

Results

- Random Forests produced the most accurate prediction of lateness for all rail lines
- Interactive code allows users to predict how late they will be to their desired station within 2.5 minutes in most cases

```
Please enter the source station: Marcus Hook
Which station do you want to know the status for?: Ridley Park
Do you want to know the status at the current time? (y/n): y
The train is predicted to be 0:01:00 late (H:MM:SS).
Do you want to quit?(y/n): n
Please enter the source station: Marcus Hook
Which station do you want to know the status for?: University City
Do you want to know the status at the current time? (y/n): y
The train is predicted to be 0:01:23 late (H:MM:SS).
Do you want to quit?(y/n):
```