**Project 1: Predicting Rail Car / Locomotive Failure for effective Preventive Maintenance**

Client: Rolling Stock Maintenance Managers to minimize maintenance cost of equipment failure.

Problem: Equipment failures occur before scheduled maintenance. How do we find patterns for the failures and include simple to understand rules to be included in preventive maintenance practices to convert them to a predictive maintenance system?

Deliverables:

* Model to predict failures before they occur minimizing false positives while preventing train disruptions.
  + Identify equipment classes prone to failures
    - Propose a method to update the maintenance plan to include the model result findings as new data is incorporated.
* Flag the least/most probable failures

**Project 2: Predicting Transit Time of Trains based on traffic patterns.**

Client: Traffic / Operations Managers to aid in scheduling of trains based on demand.

Problem: Demand is not constant so static traffic grooves/schedules cause delays. How do we find patterns for delays and include simple to understand rules to be included in scheduling practices to convert them to improve locomotive/cars utilization and minimize transit time?

Deliverables:

* Model to predict transit time given train consist (number of locomotive / cars), traffic congestion and track/speed restrictions.
  + Identify efficient schedules given a level of demand.
    - Propose a method to update the train schedule to include the model result findings as new data is incorporated.
* Flag the least/most time-consuming restrictions.

**Project 3: Predicting Fuel Consumption based on available equipment, crew and demand.**

Client: Operation Planning Managers to aid in cost effective locomotive assignment on trains and yards based on demand.

Problem: Demand is not constant so there is a dynamic assignment of locomotives based on demand. How do we find the most efficient fuel consumption and include simple to understand rules to be included in equipment assignment practices to improve fuel efficiency in locomotives?

Deliverables:

* Model to predict fuel consumption given locomotive, freight weight, track section (gradient/slope), equipment type, crew, distance, time, direction.
  + Identify fuel efficient of each type of locomotive.
    - Propose a method to update equipment assignment to include the model result findings as new data is incorporated.
* Flag the least fuel-consuming assignment decisions to alert cost differences.

How much data do you have on work orders?

Ideally you'd have train level data like all the information about it for last year say, every trip, repair, etc. Then using that plus all the work order data you could try build a model to predict types of faults in next month or Xkm say.

I'd imagine this type of data might not really exist in one place and making it might not even be possible.

In that case if you have a decent amount of work orders then you could maybe do things like shopping basket analysis on the errors you see. So think of the work orders data as like a shopping history of each train and maybe you could then see what types of errors occur together to be able to implement into repair guidelines etc.

Or you could maybe group it to monthly data to try see what types of tracks, and engines, trains etc are more likley to have what types of problems etc.

Or you could maybe try focus on work order efficiency to see if you can predict how long a specific repair might be or even maybe what the specific problem is most likely to be given everything you know as you start the repair.

Like  Reply

this otp data is cool. very similar to the airline flights data but cooler as has stations.

is this your own data? open source? keen to understand how much there is.

for ideas here i'd look into things people have done with the airline delays data.

some example links from a wquick google of "airline flight delay data project"

also much more beyond just predicting delays. you could do things like network analysis, clustering of stations based on usage etc.

would be cool to enrich this data as well with things like weather, calendar holidays, events, etc.

Like  Reply

idea 3 is a perfect out of the box ready to go application of data science. do you have a feeling by how much an advanced model like this would beat a simple calculation based on distance in terms of accuracy - that would even help you end up quantifying potential cost impacts once you have a simple benchmark to check against and beat (maybe the fancy model might not even end up being that much better then a simple calc based on distance but who knows?)

Like  Reply

<http://ddowey.github.io/cs109-Final-Project/#!index.md>  
<http://trap.ncirl.ie/1850/1/ciaranohurmoltaigh.pdf>  
<https://www.slideshare.net/HaozheWang/airline-flights-delay-predicton>  
<https://hortonworks.com/blog/data-science-apacheh-hadoop-predicting-airline-delays/>  
<https://jessesw.com/Air-Delays/>  
<http://www.ns.umich.edu/new/multimedia/videos/22592-how-big-data-could-reduce-weather-related-flight-delays>  
<http://cs229.stanford.edu/proj2012/BandyopadhyayGuerrero-PredictingFlightDelays.pdf>  
<http://e-collection.library.ethz.ch/eserv/eth:5404/eth-5404-01.pdf>