Data Science @ Ike: Homework

Introduction

This is a hypothetical exercise relevant to the Data Science position here at Ike. The purpose of the exercise is twofold:
i) Provide a flavor of the type of work that will be involved in the position
ii) Provide material that may be used for more detailed discussions in on-site visits.

Guidelines:

- Please use whatever analysis tools, references, and documentation styles you prefer.
- Please focus on describing the approach (e.g. “This is how I would approach this”) rather than arriving at a complete solution.
- We hope you have fun! Unconventional ideas and approaches are welcome!
- Please include code along with any writeup of results/conclusions (ipython notebooks are great!)

Estimating mean distance to failure

Background: High cross-winds present a danger to long-haul truck drivers. High winds have the capacity to cause tractor-trailer rollovers. These are especially pernicious because they are often difficult to anticipate and very difficult to mitigate.

Exercise: Imagine that as part of our fleet safety evaluation, we would like to estimate the mean distance to failure (similar to a mean time to failure) for the failure case of a rollover caused by crosswind. For this hypothetical example, assume that we would like to operate in central Canada along Highway 1 between Calgary and Winnipeg.

1. Using publicly available weather data, please estimate the frequency of high wind events that could potentially cause rollover conditions (frequency can be measured in hours of operation per year). Is there significant geographic, diurnal, or seasonal variation?

Some tips: i) Historical weather data may be retrieved using wget following instructions from: ftp://ftp.tor.ec.gc.ca/Pub/Get_More_Data_Plus_de_donnees/Readme.txt. It can also be retrieved as

csv downloads from a web interface at: http://climate.weather.gc.ca/historical_data/search_historic_data_e.html. Feel free to use other data sources or methods of data retrieval. ii) Not all weather stations report wind speed. Stations WINNIPEG A CS and CALGARY INT'L CS have reasonably good wind reporting as do most airports. Also note that max wind gusts are included in daily reports, but not hourly reports. To find the wind threshold that corresponds to rollover risk, feel free to use citations/references rather than performing the calculation.

2. Using publicly available analysis and data, please provide an estimate of a mean distance to failure for wind-induced rollover events for our hypothetical lane along Highway 1 between Calgary and Winnipeg. Please use existing literature to generate a rough understanding of historical incident rates, then create a statistical argument to convert this into an estimate of the mean distance to failure. (A mean time to failure estimate can be used as well.) Please include a description of your assumptions as well as the associated math.