---------- DS challenge ----------

Chicago Ridership

This DS challenge is designed to evaluate your skills and intuition in a real world data problem.
Please using the following dataset from 2002 to 2016
Dataset: https://data.cityofchicago.org/Transportation/CTA-Ridership-L-Station-Entries-Daily-Totals/5neh-572f

CTA - Ridership - 'L' Station Entries - Daily Totals | City of Chicago | Data Portal CTA - Ridership - 'L' Station Entries - Daily Totals - This list shows daily totals of ridership, by station entry, for each 'L' station dating back to 2001. Dataset shows entries at all turnstiles, combined, for each station. Daytypes are as follows: W=Weekday, A=Saturday, U=Sunday/Holiday. See attached readme file for information on how these numbers are calculated.

Load data and perform exploratory data analysis.

What are the characteristics of the data?
What are your findings?
What did you learn from the data?
The transportation department wants to improve service in the next few years. Can you build a model for them to forecast daily rides? (Please use 2017 data as testing set for evaluation) Which aspects should you consider for this model?

Please explain how you built the model and justify the choices you made.
How would you evaluate this model to ensure it would be robust for production usage?

Please submit the result in the form of runnable notebooks or scripts. A link to GitHub or other code repository would be great.
Please let us know if we need to do anything special to run your notebook (install packages, get extra data etc.)