# Capstone Project Ideas

## Walmart M5 Forecasting - Accuracy and Uncertainty

### Business Problem

Walmart needs improved point and probabilistic forecast estimates of the unit sales of various products sold in the USA by Walmart. Improved forecasts will enable Walmart to reduce lost actual or opportunity costs that occur when demand is under- or overestimated.

### Stakeholders

Makridakis Open Forecasting Center (MOFC). MOFC conducts forecasting research and provides companies with business forecast training.

Walmart is an American retail corporation that operates a large chain of stores selling discounted items and groceries.

Kaggle is an online community of data scientists and machine learning practitioners.

### Datasets

Hierarchical data from three Walmart stores that includes item level, department, product category and store details.

### Data Science Approaches

Numerical forecasts for accuracy and uncertainty will require some form of regression. To augment the regression, classification features may be helpful. Certain classification data are provided.

---

Evaluation

Results will be evaluated in a competition to provide the most accurate forecasts. Uncertainty forecasts are scored using a Weighted Scaled Pinball Loss (WSPL). Accuracy forecasts are scored by Weighted Root Mean Squared Scaled Error (RMSSE).

### Usage of Results

Successful forecasts will help to advance the theory and practice of forecasting. Such methods can be applied in various business areas, such as setting up appropriate inventory or service levels.

# Power Laws: Cold Start Energy Forecasting

https://www.drivendata.org/competitions/55/schneider-cold-start/page/110/

## Business Problem

Building energy consumption forecasts are required for energy planning policies, energy storage optimization, and for detecting wasteful anomalies. In this problem, we look at forecasting energy usage for buildings under construction for which very little energy usage data are available.

## Stakeholders

Energy utilities (electric power, gas)
Building owners and facility managers
Government

## Datasets

A .csv file of “cold start” measurements of building consumption data (watt-hours) and outside temperature. Cold start measurements are the first data collected at the beginning of the building’s energy instrumentation life.

## Data Science Approaches

Energy consumption forecasts will require regression techniques. Estimates are to be provided for hourly consumption over a single day, daily consumption for one week, and weekly consumption over two weeks.

---

## Evaluation

The results are evaluated using normalized mean absolute error. The weighting used for this metric makes each prediction equally important in the performance metric score.

## Usage of Results

This problem was part of a competition (now closed) hosted by Driven Data. I anticipate that this forecasting tool and others like it will be used as input for purposes like electric utility planning, government energy policy planning, and climate change forecasts.