Capstone 3 – Project Proposal

Background: After revising the 6 possible datasets for this project, I have decided to finally choose the [Corporación Favorita dataset](). The logic behind this selection is the following:

- It addresses a time series problem which I am particularly interested at the moment given the type of work I need to do at my new job

- With this dataset we can ask a very particular and relevant business questions and hopefully give some answers to it through actionable insights

- The whole dataset is comprised of sub-datasets that can allow us to do a multivariate time series analysis which can be interesting for learning purposes.

Problem statement: Can we accurately forecast sales of particular family of products in each store?

- **Context:** For the sake of this project, we will consider the following scenario: Corporación Favorita has asked us to help them predict how sales will evolve for certain family of products (meats, seafood, cleaning, etc.) during the upcoming weeks. Additionally, management is interested in any insights we can produce from their data and hopefully offer them useful suggestions on how to move their business forward.

- **Criteria for success:** The information we extract from this analysis will help us offer Corporación Favorita insights in relation to how much of each product they should stock up for the upcoming weeks. We will determine success if we are able to obtain a [RMSLE]() (Root Mean Squared Logarithmic Error) of less than 0.52 with our test data.

- **Scope of solution space:** We are focusing only on the data provided by Corporación Favorita. This includes 54 stores, and 4100 items, daily oil price evolution, and holidays information. We will refrain of getting data from other datasets.

- **Constraints:** We do not have detailed information about each product. Corporación Favorita has decided to keep the names of their products hidden (we are only given a item ID). This makes it difficult to offer particular insights for each product.

- **Stakeholders:** The main and unique stakeholder is Corporación Favorita

- **Data sources:** I will be using the Kaggle dataset "[Corporación Favorita Grocery Sales Forecasting]()"