Project 1 proposal

1. Team members
   1. This project will be completed by Alex Deshowitz
2. Project description
   1. This project seeks to create market level elasticity curves at the O&D market level for Southwest Airlines. Southwest Airlines serves over 650 nonstop markets each day with O&D combination possibilities ranging from 5500 – 6000. We will create curves for the top markets with O&D passengers per day each way (PDEW) being the primary filtering characteristic. We will limit the study to markets with at least 100 PDEW in order to ensure a large enough sample set from which to derive the models.
3. Features to be input to the model
   1. All data will be obtained from the DOT O&D survey and publicly available schedules. Unfortunately, Southwest Airlines has declined to allow internal booking curve data to be used for this project. Therefore, we will use publicly available sources to munge a dataset together. This data is only available at the quarterly level. However, we should be able to go back to around the year 2000 and pull together some quality insights.
   2. Response
      1. Clearly the response variable will be the percentage change in demand (O&D PDEW)
         1. I will need some help on defining this, I think it is as simple as taking the log of the response variable?
   3. Explanatory
      1. Percentage change in price
         1. I think we will have to take the log of this variable as well? Rather than calculate a change over a period be it year-over-year or quarter to quarter?
      2. One hot encoded quarters
         1. There will potentially be some seasonality in the data; therefore, we may need to test for the quarterly seasonality. Also, if holidays (such as Easter) shift between quarters, we will account for this variability in the data
         2. Could change this by making the data YE quarters, this would effectively lower the number of observations per market by 4
      3. Other Airline (OA) seats
         1. A factor that can impact elasticities is the presence and/or change in the number of seats that other airlines offer in a market.
      4. Southwest Airlines seats in the market
         1. The number of seats that Southwest offers in the market will also be directly related to the demand in a market. In some markets, adding seats may have a stimulative impact whereas in other markets, adding seats may result in no additional demand.
      5. Ultra-low-cost carrier (ULCC) presence
         1. The presence of ULCC carriers in a market tends to bring prices down due to the impact of the “a la carte” model being implemented in the market.
      6. Subsequent models will incorporate additional data into the process, but these variables should provide a good starting point as a proof of concept.
4. Data Sources
   1. O&D data
      1. DOT DB1B data obtained from the DOT website
      2. Quarterly data sample provided to the government by all US passenger carriers
   2. Schedule data
      1. Publicly available SSIM files provided by airlines
      2. Each week, all domestic carries provide SSIM files to reservation systems and data aggregators so that schedules may be published to online travel sites and be placed available for sale. We will leverage this data for the model.
5. Outcomes
   1. The overall outcome should be a better understanding of the elasticity of each type of market in the Southwest Network. When fully built out, this POC could potentially provide planners with a better idea of how different variables may impact the demand and overall strategy for the markets.
   2. Eventually, it would be pretty neat if Southwest were able to cluster these markets into groups with distinct characteristics and have more globalized strategies incorporated into the business management process.
   3. Measurement
      1. We will use an accuracy based approach as that seems to be the best way to convey results and model reasoning to executives.