ISYE-6501
Fall 2024

Course Project

**The websites https://www.sas.com/en_us/customers.html, https://www.ibm.com/case-studies/search?search, and https://www.informs.org/Impact/O.R.-Analytics-Success_Stories (among others) contain brief overviews of some major Analytics success stories. In this course project, your job is to think carefully about what analytics models and data might have been required.**

(1) **Browse the short overviews of the projects. Read a bunch of them - they're really interesting. But don't try to read them all unless you have a lot of spare time; there are lots!**
(2) **Pick a project for which you think at least three different Analytics models might have been combined to create the solution.**
(3) **Think carefully and critically about what models might be used to create the solution, how they would be combined, what specific data might be needed to use the models, how it might be collected, and how often it might need to be refreshed and the models re-run. DO NOT find a description online (or elsewhere) of what the company or organization actually did. I want this project to be about your ideas, not about reading what someone else did.**
(4) **Write a short report describing your answers to (3).**

I decided to go with the Audi-UK case study because I really like cars and used to own an Audi. Please find below the link to the case study.
https://www.ibm.com/case-studies/audi-uk

**Analytical Models:**

- **Linear Regression**
  - Audi would use a predictive model to forecast customer behavior. For example, which users are most likely to book a test drive or make a purchase can be found out with linear regression.
- **SARIMA**
  - Audi would use a descriptive model to better understand historical performance. For example, SARIMA can show which is the busiest time of the year and which car models are the most popular using seasonality trends.
- **Optimization model**
  - With the help of the Predictive and Prescriptive analysis, we can use k-means clustering to group the different types of customers to cater their ads and promotions they get.

**How were they combined:**

The Linear regression model can be used to determine which customers are most likely to book a test drive based on their different attributes. The Y (response) would be a range from 0-1 with 1 being the most likely to test drive and 0 being very unlikely to test drive. This would help us determine which customers are more likely to not only test drive but also have a higher chance of purchasing or leasing a

vehicle. The SARIMA model can be used to find trends in seasonality and sales numbers. The SARIMA model would probably show some type of seasonality such as most cars are sold near the end of the month and the end of the year due to the customer knowing dealerships need to meet their quotas. Now with the help of the linear regression model finding the test-driving customers, and the help of SARIMA to determine trends in sales and seasonality, they are able to use some type of optimization model such as Non Linear Programming to find the what is the best time of the year and the best type of customers to show catered promotions and advertisements.

**Data Requirements:**

- **Linear Regression Data:**
    - Demographics
        - Age
        - Income
        - Location
        - Employment
    - Customer History (# of visits to website within last month, past purchases)
        - # of visits to the website within the last month?
        - Purchased a car within the last 20 years?
- **SARIMA Data:**
    - Sales History
        - Which month shows the highest sales?
        - Which car model shows the highest sales?
    - Trend History
        - Have sedans or SUVs had rising sales compared to the last decade?

- **Optimization Model:**
    - Data from both the Linear Regression Model and the SARIMA Model.

**Data Collection:**

- **Linear Regression Data:**
    - The patient demographic data can be collected with the use of a small survey that the customers are emailed or are asked to fill out when they visit the website.
    - Getting customers to fill out surveys can be a real pain so in order to get a higher response rate, we can incorporate some type of reward such as a very large car raffle.
- **SARIMA Data:**
    - The past sales history can be retrieved from the informatics team at Audi. They must have an IT team that keeps track of all sales in a very large data warehouse.
    - The trend data can be retrieved from previous sales within the last decade to see which types of cars are becoming more popular.
- **Optimization Model:**
    - The data used for this is taken from the results of the two other models.

**Refresh Rate:**

The data refresh most likely would be done depending on how often the company wants to change their advertisements and promotion campaigns. If the promotion campaigns change quarterly then the data should be refreshed before the campaign data is presented. When presenting data periodically, it is always very important to make sure that the data is updated to reflect the most recent time period.