The following projects are examples of what I was working on during my master's degree in Data Science at Texas A&M University

Project Title	Advisory System for Fragrance Ingredients using Search Algorithms
Field / Area	Artificial Intelligence and Machine Learning
Description	Problem Statement & Motivation: Since I'm a certified perfumer, I want to address two real-life problems that perfumers face constantly which are: 1. No accessible database of fragrance ingredients 2. Replace the traditional way, trial & error, of choosing suitable fragrance ingredients My approach: The idea is to build a trusted database and that will feed into an advisory system to recommend which fragrance ingredients to use based on perfumer's input using searching algorithm. My approach is divided in two parts: The first part: Building a database: Having a trustworthy website that provide detailed information about fragrance ingredients is really challenging. My approach is to write a code that scrap the data and search for specific information from multiples well-known websites and store them in an excel file. To do this, I will use Scrapy framework in python which is large scale web scraping tool to efficiently extract data from websites. Then, I will do the data cleaning process and preparation for the next part. The second part: Design fragrance ingredients' advisory system: The idea is to design an advisory system based on searching algorithms that helps perfumers to decide which ingredients to use in order to have creative blends. The input of the system will be descriptive words of what the perfumer wants to achieve as a final scent. For example, flowery, woody, and musky. Then the system uses searching algorithms to scan the database looking for best ingredients that match what the perfumer is asking for. The output will be some recommended fragrance ingredients including essential oils and aroma chemicals.
Programing language	**Python**

Project Title	Texas Natural Fire Analysis and Visualization
Field / Area	Spatial Statistics
Description	Introduction: Wildfires occur often around the United States, especially in hotter and drier climates and can cost either millions of dollars in damages or catastrophic natural damages to our environment. Many local governments are tasked with trying to figure out ways to dampen these financial and environmental losses by assessing large datasets over long periods of time and making decisions that provide relief to the overall negative effects that wildfires can cause. Problem Background: The problem that I'm trying to solve is how can I visualize spatially the spread of wildfires in the US and specifically the state of Texas. Also, I'm planning to design model to find either trends or relationships among the wildfire locations. These trends that we find can ultimately lead to either an infographic or visualization that shows local government officials where action needs to be taken. Goal / Techniques Used: My goal with this project and analysis is to be able to use interactive visualizations and predict the time spent on putting out individual wildfires based on their location using a Gaussian Process Spatial Regression Model and advanced visualization techniques like R-shiny app.
Programing language	**R using R-studio and R-shinny App**

Project Title	Stock Market Clustering and Predictions
Field / Area	Applied Data Analytics
Description	There is too much unpredictability present in stock market data which makes it difficult to predict with regular linear regression techniques. The goal of this project is to do classification and predictive modeling of Stock Data to provide recommendations on which companies to invest in for a specific period. We apply Fast Fourier transform algorithm to classify companies based on the seasonality of their time-series data. Then, we will use ARIMA (autoregressive integrated moving average) model to forecast the future performance of these companies. At the end, we will design an interactive website using R-shinny App to easily visualize the companies' forecasted results.
Programing language	**R using R-studio and R-shinny App**
Implementation	https://alnuaimi107.shinyapps.io/stock-market-pnc/

Project Title	Formula 1 Analysis & Prediction
Field / Area	Regression Analysis
Description	The idea is to analyze Formula-1 dataset and trying to fit a statistical model that can accurately predict specific parameter. In our case, we use the following parameters : fastestLapTime, circuitId, driverId, laps to estimate the positionOrder that is which driver will most probably win the race. The two models used for data fitting are Generalized linear model (GLM), a regular linear model and Multinomial logistic regression model.
Programing language	**Python & R using R-studio**

Project Title	Trekommender (The Transcendent Social Travel System)
Field / Area	Data Mining and Machine Learning
Description	Recommender systems are becoming essential in many industries and, hence, have received more attention in recent years. Applying what we learned in class with other data mining and user interface algorithms, our team successfully deployed the Trekomender website that enables users to make their decision of where to go much easier. the website not only recommends a destination but also gives a high-level description and highlights what are the most popular sights and activities in that destination. It also suggests what is the best time to visit. Imagine all this useful information will be available to you once you share your preferences in traveling!!!
Programing language	**Python , Spark**