Capstone Project Ideas

[Capstone Project Idea 1: Analyze historical demand, and forecast sales. 2](#_Toc495426972)

[What is the problem you want to solve? 2](#_Toc495426973)

[Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise? 2](#_Toc495426974)

[What data are you going to use for this? How will you acquire this data? 2](#_Toc495426975)

[In brief, outline your approach to solving this problem (knowing that this might change later). 3](#_Toc495426976)

[What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck. 3](#_Toc495426977)

[Capstone Project Idea 2: Analyze historical stock price data and predict closing price 4](#_Toc495426978)

[What is the problem you want to solve? 4](#_Toc495426979)

[Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise? 4](#_Toc495426980)

[What data are you going to use for this? How will you acquire this data? 4](#_Toc495426981)

[In brief, outline your approach to solving this problem (knowing that this might change later). 4](#_Toc495426982)

[What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck. 4](#_Toc495426983)

[Capstone Project Idea 3: Measure college performance. 5](#_Toc495426984)

[What is the problem you want to solve? 5](#_Toc495426985)

[Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise? 5](#_Toc495426986)

[What data are you going to use for this? How will you acquire this data? 5](#_Toc495426987)

[In brief, outline your approach to solving this problem (knowing that this might change later). 5](#_Toc495426988)

[What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck. 5](#_Toc495426989)

# Capstone Project Idea 1: Analyze historical demand, and forecast sales.

### What is the problem you want to solve?

Predict the sales based on historical demand and holiday markdown events for 45 stores located in different regions. Each store contains many departments and there is a need to project the sales for each department in each store. Markdowns affect sales, and the challenge is to predict which departments are affected and the extent of the impact.

### Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

This is a past challenge problem taken from Kaggle. This challenge was posted by Walmart Stores. Based on the analysis, Walmart will be able to make better inventory handling decisions based on predicted sales forecast.

### What data are you going to use for this? How will you acquire this data?

**The stores information, training and test data and features data is available from Kaggle for this project.**

#### ****stores.csv****

This file contains anonymized information about the 45 stores, indicating the type and size of store.

#### train.csv

This is the historical training data, which covers to 2010-02-05 to 2012-11-01. Within this file you will find the following fields:

* Store - the store number
* Dept - the department number
* Date - the week
* Weekly\_Sales -  sales for the given department in the given store
* IsHoliday - whether the week is a special holiday week

#### test.csv

This file is identical to train.csv, except we have withheld the weekly sales. You must predict the sales for each triplet of store, department, and date in this file.

#### features.csv

This file contains additional data related to the store, department, and regional activity for the given dates. It contains the following fields:

* Store - the store number
* Date - the week
* Temperature - average temperature in the region
* Fuel\_Price - cost of fuel in the region
* MarkDown1-5 - anonymized data related to promotional markdowns that Walmart is running. MarkDown data is only available after Nov 2011, and is not available for all stores all the time. Any missing value is marked with an NA.
* CPI - the consumer price index
* Unemployment - the unemployment rate
* IsHoliday - whether the week is a special holiday week

For convenience, the four holidays fall within the following weeks in the dataset (not all holidays are in the data):

Super Bowl: 12-Feb-10, 11-Feb-11, 10-Feb-12, 8-Feb-13  
Labor Day: 10-Sep-10, 9-Sep-11, 7-Sep-12, 6-Sep-13  
Thanksgiving: 26-Nov-10, 25-Nov-11, 23-Nov-12, 29-Nov-13  
Christmas: 31-Dec-10, 30-Dec-11, 28-Dec-12, 27-Dec-13

### In brief, outline your approach to solving this problem (knowing that this might change later).

1. Using exploratory data analysis, study the dataset, look for any outliers or missing data and clean up as needed.
2. Identify which statistical model or algorithm can be used to predict the target variable ‘sales’.
3. Apply the algorithm on the train and test data set, and compare the predicted results.
4. Review the metrics/results and ensure that the chosen model/algorithm best fits the analysis.

### What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

Code, analysis report and presentation slide.

# Capstone Project Idea 2: Analyze historical stock price data and predict closing price

### What is the problem you want to solve?

 With the historical stock prices and adjustment data, predict the closing price of the current day.

### Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

My clients are those who are looking to buy/sell the stocks.

The end of day stock price dataset of different companies (AAPL, HD, VZ) are retrieved from Quandl. Predictions are made on the closing day stock price based on their historic stock prices and dividends/adjustments. This will help my clients to buy/sell these stocks accordingly.

### What data are you going to use for this? How will you acquire this data?

Historical, end of day US stock prices, dividends, adjustments and splits for all publicly traded stocks is available in Quandl. This data is updated daily. I am planning to use Apple’s, HomeDepot’s and Verizon’s data for analysis.

EOD\_AAPL.csv, EOD-HD, EOD-VZ datasets that are available in Quandl will be used.

### In brief, outline your approach to solving this problem (knowing that this might change later).

1. Using exploratory data analysis, study the dataset, look for any outliers or missing data and clean up as needed.
2. Identify which statistical model or algorithm can be used to predict the target variable.
3. Apply the algorithm on the train and test data set, and compare the predicted results.
4. Review the metrics/results and ensure that the chosen model/algorithm best fits the analysis.

### What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

Code and presentation slide.

# Capstone Project Idea 3: Measure college performance.

### What is the problem you want to solve?

Provide information of college performance so students can evaluate their college choices.

### Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

The analysis will guide the students to search for colleges that are a good fit for them. They will be able to make informed decisions on which degree to pursue and at which institution.

### What data are you going to use for this? How will you acquire this data?

CollegeScorecard Raw data is available in the catalog.data.gov that can be used for analysis.

<https://catalog.data.gov/dataset/college-scorecard>

### In brief, outline your approach to solving this problem (knowing that this might change later).

1. Using exploratory data analysis, study the dataset, look for any outliers or missing data and clean up as needed.
2. Identify which statistical model or algorithm can be used to predict the target variable.
3. Apply the algorithm on the train and test data set, and compare the predicted results.
4. Review the metrics/results and ensure that the chosen model/algorithm best fits the analysis.

### What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

Code and presentation slide.