Solution Design and Data Collection:

The goal of the project as stated in the project statement is to come up with a model to predict the price of a certain stock based on various indexes and ETFs in which this stock is a component. We have selected the US Banking Services Industry as our target industry and we will be analyzing JP Morgan (NYSE: **JPM**) to perform our analysis. The goal is to predict based on the performance of various indices whether the stock price will go up or down during the current day.

We will be having 4 different datasets.

1- Target Stock: JP Morgan Stock Dataset, a daily stock dataset with variables such as **High, Low, Close** for a period of 1 year. Available on Yahoo Finance API for R in the R Package Quantmod.

2- Index 1: Dow Jones Bank Index (**DJUSBK**) is a popular benchmark index for top banking companies in the USA. JPM is a component of this index and has a weightage of 4.57%. The dataset for this index will also be based on 1-year time series data with **High, Low and Close** for each day. Data for this is available from various sources such as Quantmod or Investing.com.

3- Index 2: Financial Select Sector SPDR ETF (**XLF**) is a prominent exchange traded fund comprising of various companies in the financial services sector in the USA. JPM is a component of this ETF and carries a weightage of 11.49%. The dataset for this index will also be based on 1-year time series data with **High, Low and Close** for each day. Data for this is available from various sources such as Quantmod or Investing.com.

4- Index 3: KBW Bank Index (BKX) is a banking industry benchmark index comprising of 22 banking firms. JPM is a component of this index and has a volume of 19.88 MM. The dataset for this index will also be based on 1-year time series data with **High, Low and Close** for each day. Data for this is available from various sources such as Quantmod or Investing.com.

After obtaining the data from each individual dataset, we need to perform a few adjustment operations to get price-based weighting as each component stock of the index has a different price. Hence, increase coherence, we multiply volume with price of the stock on that day to generate actual percentage weight value of the stock for that day.

We then plan to use ARIMA (Auto-Regressive Integrated Moving Average) to forecast our stock price by training the ARIMA models on our benchmark indices. We will be optimizing our algorithm along the way and will be experimenting with various ML models like neural networks as well. For neural networks we will be using TensorFlow for R.

Snapshots of Datasets.

KBW Nasdaq Bank Index (^BKX) Add to watchlist
Nasdaq GIDS - Nasdaq GIDS Delayed Price. Currency in USD

112.92 +0.39 (+0.35%)
At close: 5:15PM EDT

Summary | Chart | Conversations | Options | Components | Historical Data

Time Period: Mar 16, 2017 - Mar 16, 2018 Show: Historical Prices Frequency: Daily Apply

Currency in USD Download Data

Date	Open	High	Low	Close*	Adj Close**	Volume
Mar 16, 2018	112.7600	114.0600	112.6700	112.9200	112.9200	-
Mar 15, 2018	112.9400	113.1100	112.0500	112.5300	112.5300	-
Mar 14, 2018	114.5000	114.5400	112.2000	112.4900	112.4900	-
Mar 13, 2018	115.9800	116.0700	113.8800	114.1800	114.1800	-
Mar 12, 2018	116.1000	116.5300	115.3400	115.5200	115.5200	-
Mar 09, 2018	114.4400	116.1000	114.3000	116.1000	116.1000	-
Mar 08, 2018	113.9900	114.3100	112.3700	113.6000	113.6000	-
Mar 07, 2018	112.7500	114.0000	112.3000	113.7700	113.7700	-
Mar 06, 2018	113.6000	114.0600	112.4400	113.7700	113.7700	-

KBW Bank Index (BKX)

DJ Banks 495.40 +0.35 (+0.07%)

Dow Jones Banks Historical Data

Time Frame: Daily Download Data 02/16/2018 - 03/16/2018

Date	Price	Open	High	Low	Vol.	Change %
Mar 16, 2018	495.40	495.33	501.07	495.27	435.60M	0.07%
Mar 15, 2018	495.05	496.56	497.14	492.36	159.27M	0.13%
Mar 14, 2018	494.41	502.59	502.85	493.21	199.01M	-1.35%
Mar 13, 2018	501.16	507.87	509.39	499.89	200.26M	-1.13%
Mar 12, 2018	506.88	507.83	510.66	505.57	193.32M	-0.21%
Mar 09, 2018	507.97	501.75	508.05	500.04	212.09M	2.24%
Mar 08, 2018	496.86	498.26	499.37	491.77	168.50M	-0.09%
Mar 07, 2018	497.31	493.06	498.32	491.11	198.02M	-0.11%
Mar 06, 2018	497.85	498.51	499.50	493.29	194.12M	0.20%
Mar 05, 2018	496.86	487.29	499.70	484.49	217.17M	1.18%

DJ Banks Index (DJUSBK)

Financial Select Sector SPDR ETF (XLF) Add to watchlist
NYSEArca - NYSEArca Delayed Price. Currency in USD

28.87 +0.02 (+0.09%) **28.89** +0.02 (0.07%)
At close: 4:00PM EDT After hours: 7:58PM EDT

Summary | Chart | Conversations | Profile | Options | Holdings | Historical Data | Performance | Risk

Time Period: Mar 16, 2017 - Mar 16, 2018 Show: Historical Prices Frequency: Daily Apply

Currency in USD Download Data

Date	Open	High	Low	Close*	Adj Close**	Volume
Mar 16, 2018	28.91	29.15	28.87	28.87	28.87	61,011,000
Mar 16, 2018				**0.105** Dividend		
Mar 15, 2018	29.05	29.11	28.86	28.95	28.85	47,451,600
Mar 14, 2018	29.40	29.40	28.85	28.94	28.84	57,278,900
Mar 13, 2018	29.74	29.77	29.20	29.27	29.16	47,049,400
Mar 12, 2018	29.70	29.81	29.55	29.60	29.49	51,784,500
Mar 09, 2018	29.31	29.72	29.19	29.70	29.59	54,595,700

Financial Select Sector SPDR ETF (XLF)