

# *Perspicacious4 Presentation*

*For Morgan Stanley – Columbia University 2nd Annual Summer Data Challenge*

*By Kinjal Saxena, Monish Jain, Niharika Prasad, and Nitin Sharma*



*TRANSCENDING DISCIPLINES, TRANSFORMING LIVES*



# Objective and Methodology

**Objective** of this project to apply AI/ML techniques on public data to quantitatively and qualitatively analyze three securities: **American Airlines (AAL), Disney (DIS), and PepsiCo (PEP)**

**Component 1:** Which security is over, under or neutral performer relative to **S&P 500** Index?

**Component 2:** Detailed Data Analysis to tie results back to Component 1.

**Methodologies** that are used to decide relative performance against S&P 500 index:

**Component 1:** Normalized price trend, CAGR, Correlation with S&P 500

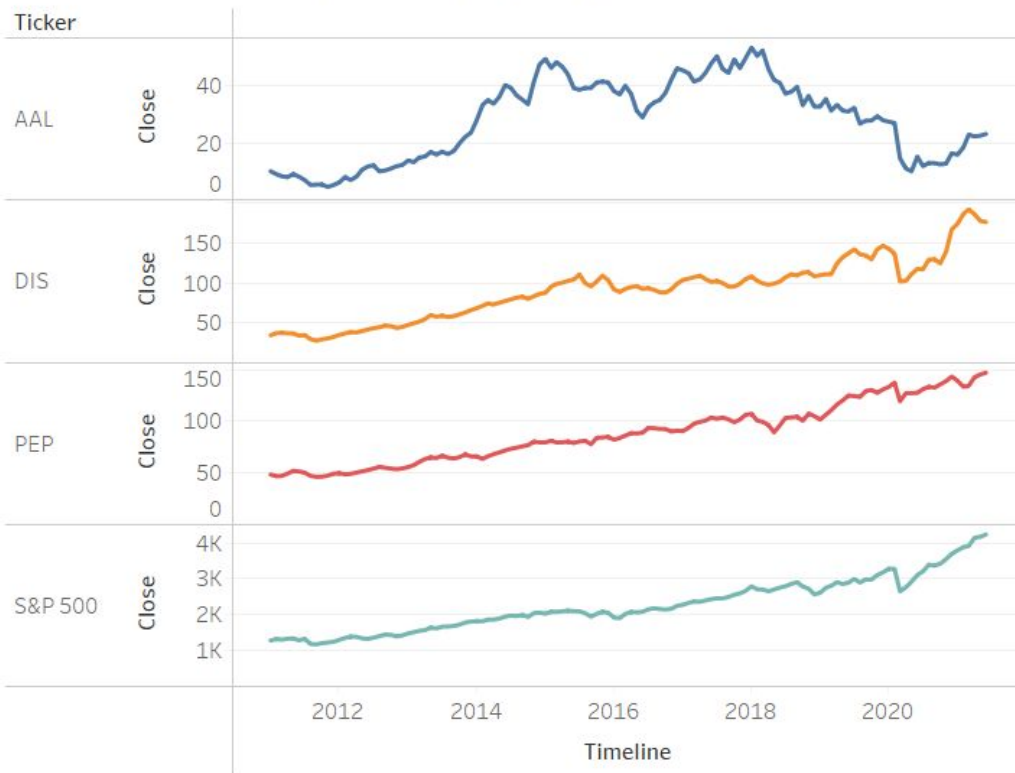
**Component 2:** Analysis using structured and unstructured data.

- **Structured Data -**
  - Linear Regression using statistical, volume, volatility, overlay and momentum indicators and autoregression factors.
- **Unstructured Data -**
  - Sentiment analysis on twitter data from major media twitter accounts.
  - Correlation between average sentiment and percentage price change.

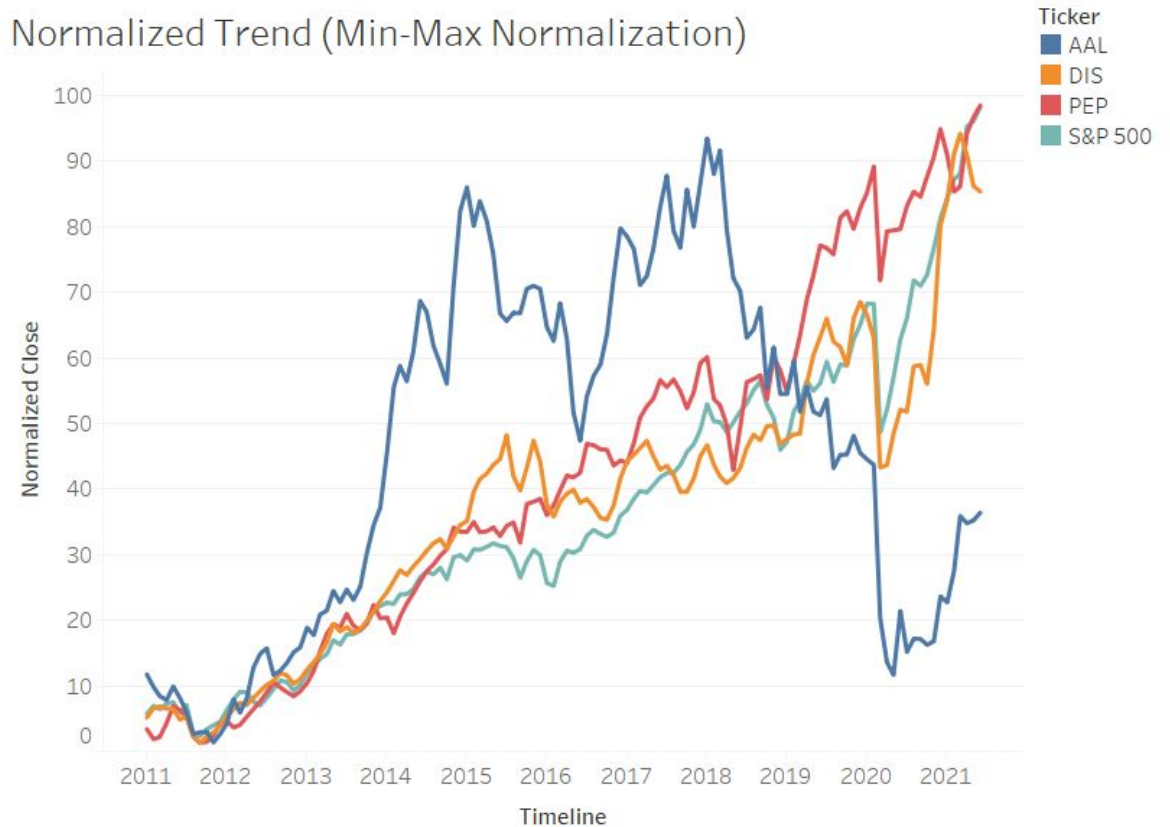
# Executive Summary of Findings

Disney (DIS) and PepsiCo (PEP) follow market trend (S&P 500). American Airlines (AAL) is more volatile and does not follow market trend, thus AAL has higher unsystematic risk.

Individual Trends (2011 - Q2 2021)



Normalized Trend (Min-Max Normalization)



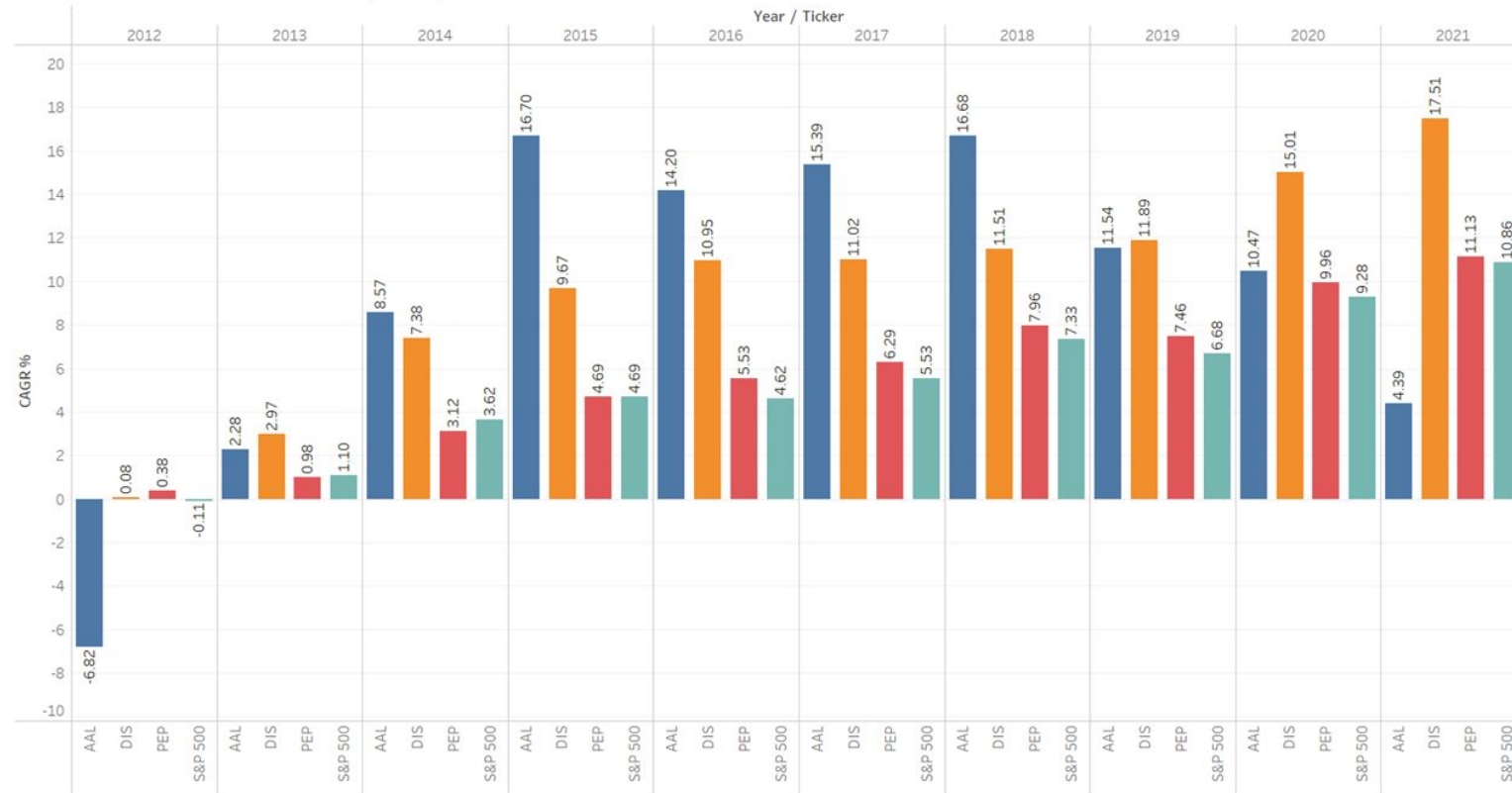
# Executive Summary of Findings

Neutral Performer compared to S&P 500 : PepsiCo (PEP)

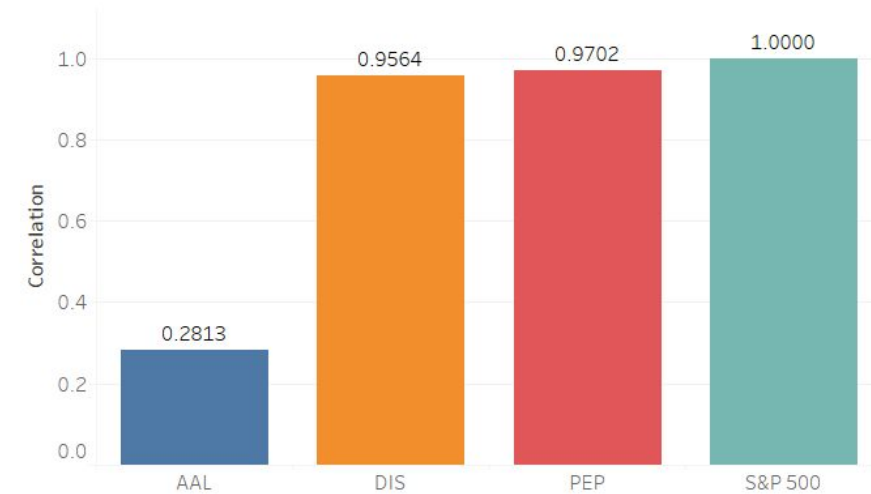
Overperformer compared to S&P 500 : Disney (DIS)

Underperformer compared to S&P 500 : American Airlines (AAL)

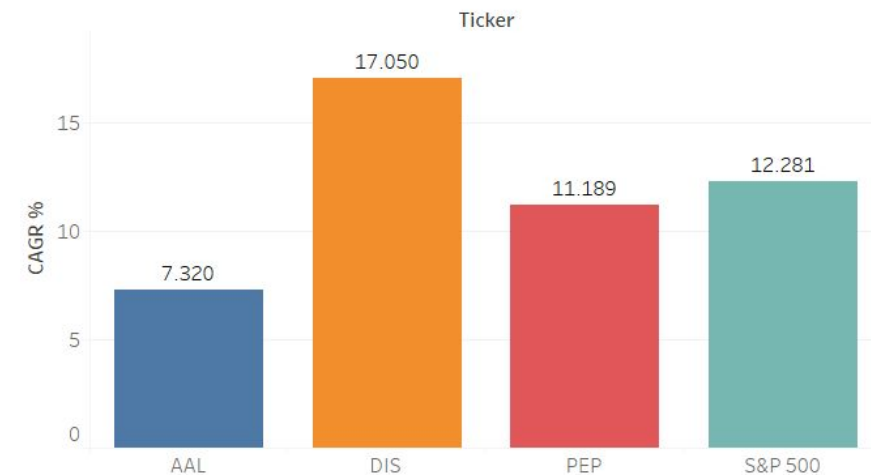
Compound Annual Growth Rate (CAGR) from 2011



Correlation with S&P 500



CAGR (Compound Annual Growth Rate)



# Detailed Data Analysis

The technical data analysis consists of following two components:

## Phase 1 - Analyzing Structured Data

Using feature engineering and factor analysis coupled with autoregression factors we found significant influencing quantitative factors and built a relatively simple linear regression model to find correlation between stock price and and impart greater interpretability to price fluctuations. We discarded data from 2020 in this analysis as it was an outlier year and hence would hamper the model's effectiveness.

Basic indicators	Autoregression factors	Momentum Indicators
Volume	Close Price on day minus 1	Rate of Change
Standard Deviation	Close Price on day minus 5	Relative Strength Index
Beta	Close Price on day minus 10	Stochastic KD
Residual Volatility	Close Price on day minus 63	Chande Momentum Oscillator
Typical Price	Close Price on day minus 252	Plus Directional Indicator
Volume & Volatility Indicators	Overlay Indicators	Minus Directional Indicator
True Range	Bollinger Bands	Average Directional Movement Index
Average True Range	Simple Moving Average	
Normalized Average True Range	Exponential Moving Average	

Indicators and their usage in the model explained in detail in appendix

# Detailed Data Analysis

## Phase 2 - Analyzing Unstructured Data

For Sentiment Analysis, we used Princeton University's Wordnet (<https://wordnet.princeton.edu>) lexical database to compute sentiment polarity and subjectivity of each word and computed overall sentiment of a sentence by calculating mean sentiment of all its words.

To begin with, we considered all tweets related to the companies under consideration irrespective of authors. This led to a plethora of irrelevant information from not so significant authors, thus creating a lot of noisy data and hampering accurate sentiment analysis. To eliminate this, we filtered tweets from the following media twitter accounts: *business*, *WSJ*, *TheEconomist*, *Forbes*, *Reuters*, *BusinessInsider*, *FT*, *CNBC*, *markets*, *YahooFinance*, *nytimes*, *HarvardBiz*, *ReutersBiz*, *Nasdaq*, *DowJones*, *NYSE*, *FoxBusiness*.

We calculated correlation between average daily stock price percentage change and average sentiment over previous quarter to determine the security's unsystematic risk and thus, corroborate its performance compared to S&P 500.



# Detailed Data Analysis - Structured Data

Top quantitative price factors which influence stock price of the securities with their regression coefficients.

## American Airlines (AAL)

R Squared = 0.98

features	regression coeff
prev_day	12.894859
EMA_45	6.723063
EMA_252	-3.701661
EMA_10	-3.335721
std_252	-2.556084
residual_volatility	2.464469
STOCH_D_252	2.130326
BBAND_UP_252	1.078323
BBAND_DOWN_45	-1.034256
STOCHF_D_252	0.973703
ATR_45	0.683454

## PepsiCo (PEP)

R Squared = 0.99

features	regression coeff
prev_day	8.376258
EMA_252	-6.200715
EMA_45	-4.273285
BBAND_UP_252	3.266580
EMA_10	2.772269
Typical Price	2.481494
SMA_252	2.328164
BBAND_252	2.328164
NATR_252	-2.130844
std_252	1.307548
BBAND_DOWN_252	1.243104

## Disney (DIS)

R Squared = 0.97

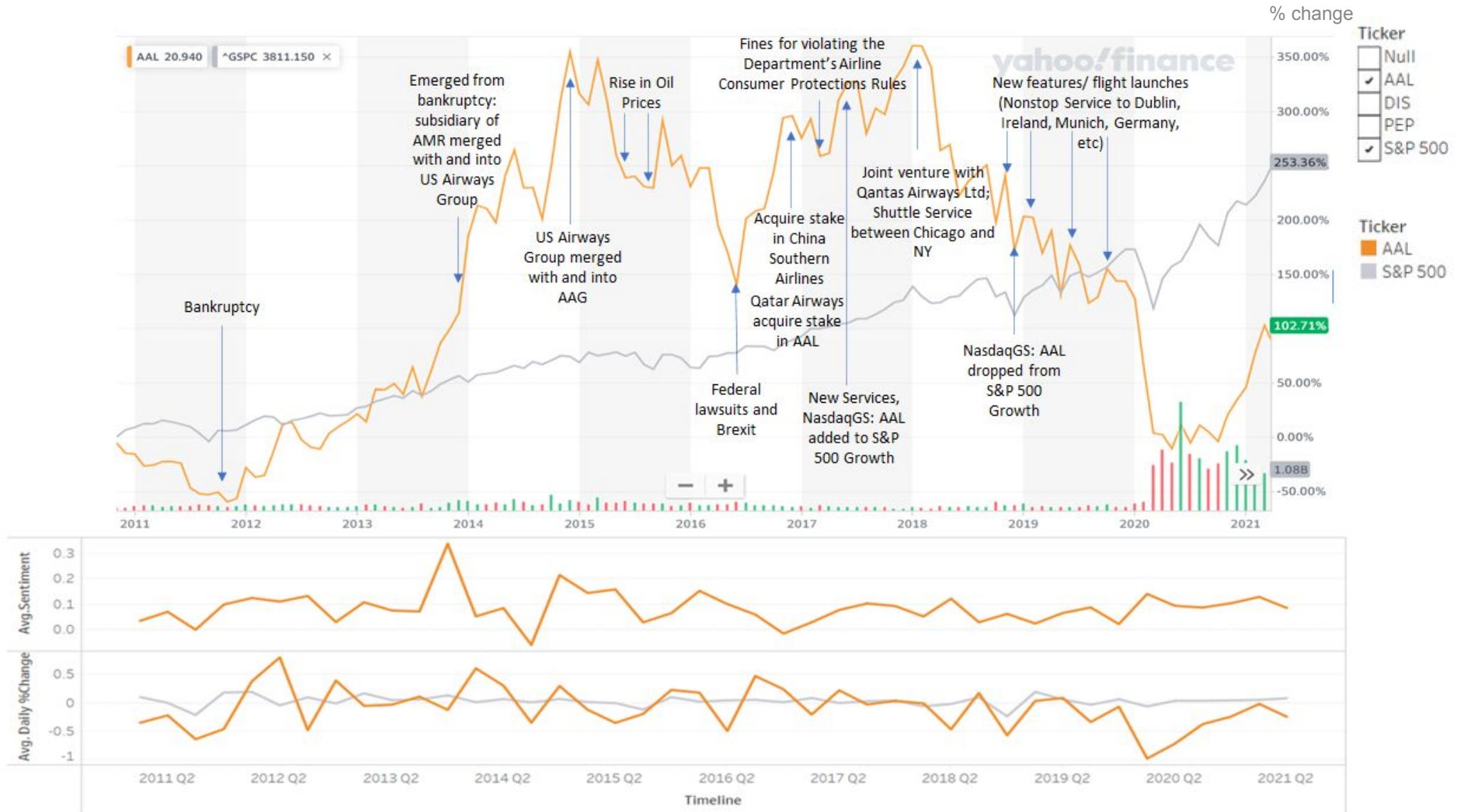
features	regression coeff
EMA_252	-22.598903
prev_day	15.132178
EMA_10	-11.046354
EMA_45	10.314908
BBAND_UP_252	5.702289
Typical Price	5.436500
BBAND_252	4.500578
SMA_252	4.500578
NATR_252	-3.100915
BBAND_DOWN_252	3.041391
ATR_252	2.458553

- Residual volatility has a greater influence for AAL and it also has some dependence on yearly stochastic oscillator bands indicating that AAL is a momentum stock and has a higher unsystematic risk, thus deviating from market trend.
- Both DIS and PEP have a positive coefficient for Typical Price indicating that these are relatively stable securities and a higher value for DIS means that on average DIS grows faster than PepsiCo, thus corroborate our analysis in component 1.

# Detailed Data Analysis & Visualization - Unstructured Data

## AMERICAN AIRLINES

**American Airlines** specific news on average have a decent influence on AAL stock price fluctuation. The daily % change follows people's sentiments on a quarter to quarter basis with a correlation of 0.25 which explains that AAL is more volatile than S&P 500.

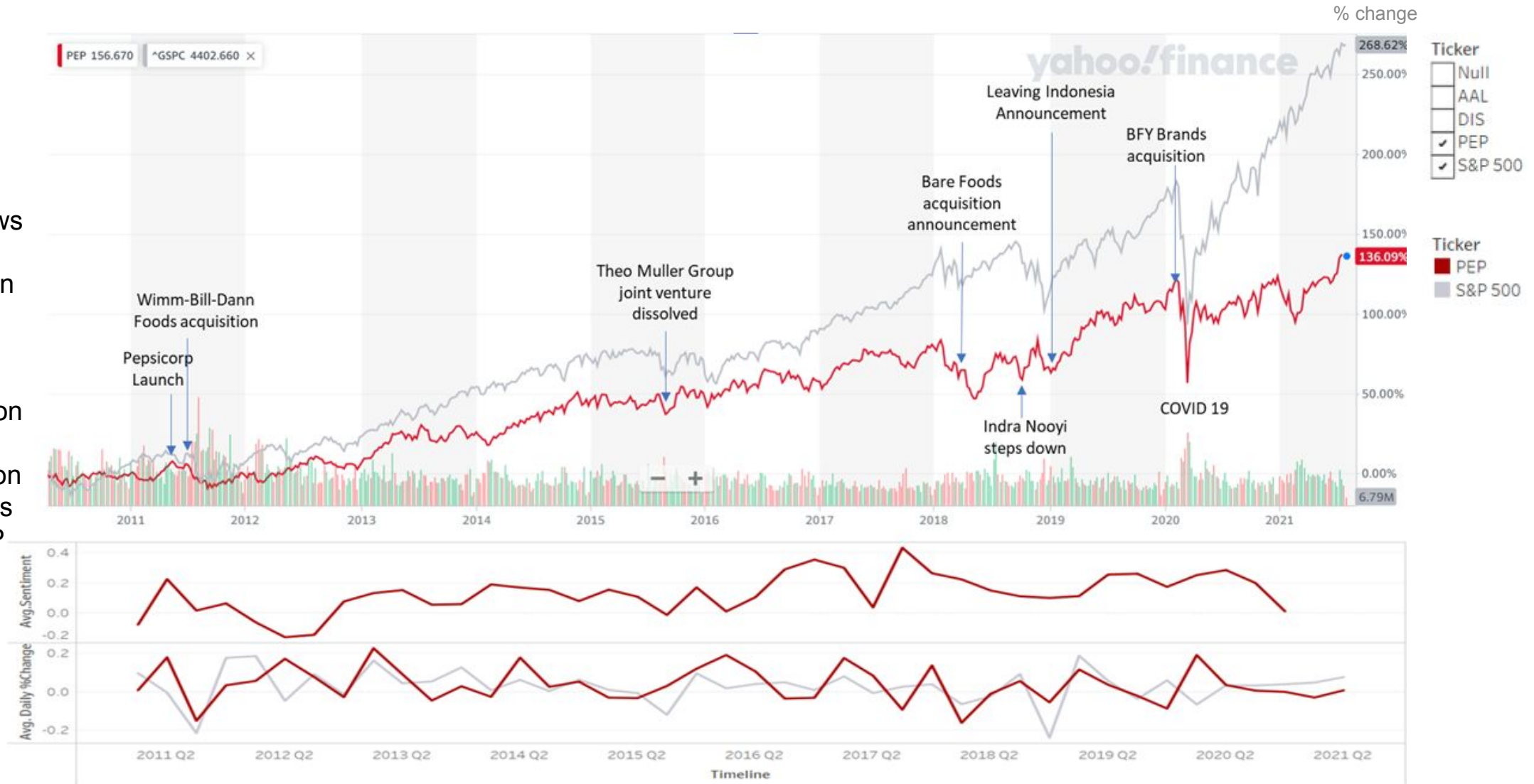




# Detailed Data Analysis & Visualization - Unstructured Data

## PEPSICO

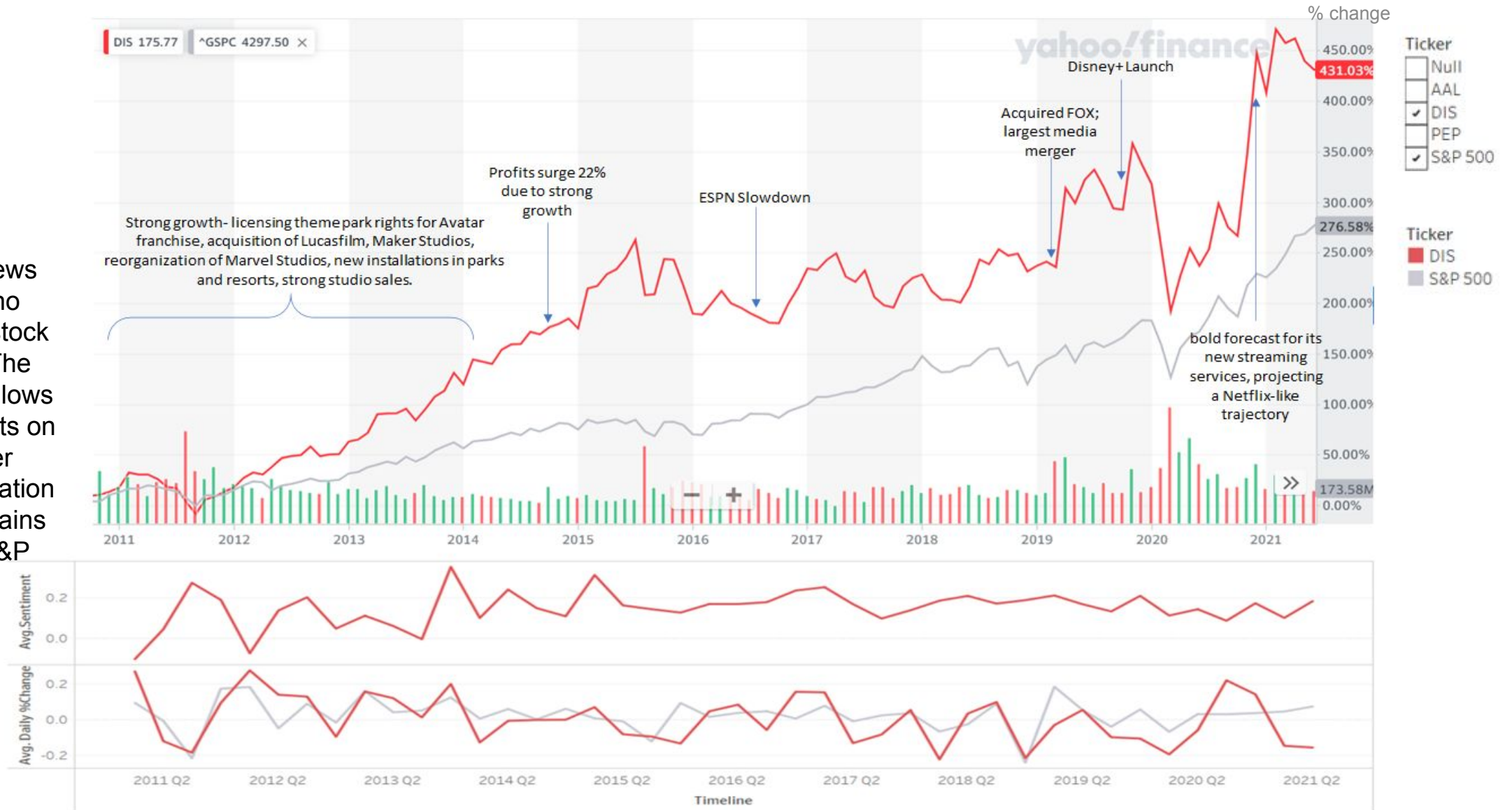
**PepsiCo** specific news on average have a negligible influence on PEP stock price fluctuation. The daily % change follows people's sentiments on a quarter to quarter basis with a correlation of 0.07 which explains that PEP follows S&P 500.



# Detailed Data Analysis & Visualization - Unstructured Data

## DISNEY

**Disney** specific news on average have no influence on DIS stock price fluctuation. The daily % change follows people's sentiments on a quarter to quarter basis with a correlation of 0.01 which explains that DIS follows S&P 500.



# Recommendations

## American Airlines will bounce back

- Highly volatile and risky (beta = 1.7) but outperformed S&P 500 by a huge margin consistently between 2014 and 2018.
- Driven by momentum. The decline in price since 2018 was predominantly due to external factors (Oil, Boeing, global demand) and nothing to do with the company's internal operations, policies and decisions.
- The drop has inadvertently created a healthy correction and made room for a bigger comeback. Thus, American Airlines can be expected to bounce back soon. However, an investor with low risk appetite should not be advised to invest in American Airlines.

## PepsiCo

- Lowest risk (beta = 0.6) among all three securities considered as well as S&P 500 and its returns match S&P 500 returns.
- New launches to take advantage of latest consumption trends and economic recovery have led to expectations of healthy growth in revenue and earnings in upcoming quarters. It is a very attractive buy for the low risk and long term investor.

## Disney has a solid outlook

- Clear long term winner. With things getting back to normal post pandemic, Disney's parks, experiences and studio arms are set to show strong growth and coupled with Disney's expected growth in the video-streaming space, Disney has a solid outlook.
- The company has been posting a Forward PE ratio upwards of 200 over the last quarter which indicates strong investor confidence in future growth.
- It's stock beta = 1 shows that on average it exactly follows S&P 500 in terms of price fluctuations. Thus, over a longer period of time it is stable, attractive and outperforming.



# APPENDIX - Basic, Volume & Volatility Indicators

Basic indicators	Feature name used in regression model	Definition
Volume	Volume	Number of shares bought and sold on given day.
Standard deviation	std_10 (trailing 2 weeks) std_45 (trailing 2 months) std_252 (trailing year)	Extent of deviation of stock price
Beta	beta	Measure of the volatility, or systematic risk, of a security or portfolio in comparison to the market as a whole
Residual Volatility	residual_volatility	Measure of unsystematic risk (risk specific to underlying security)
Typical price	Typical price	Arithmetic average of the high, low, and closing prices for a given period.

Volume & Volatility Indicators	Feature name used in regression model	Definition
True Range	TR	Maximum of current high less the current low, the absolute value of the current high less the previous close, and the absolute value of the current low less the previous close.
Average True Range	ATR_10 (trailing 2 weeks) ATR_45 (trailing 2 months) ATR_252 (trailing year)	Moving average of True Range

# APPENDIX - Overlap Indicators

Overlap Indicators	Feature name used in regression model	Definition
Bollinger bands	BBAND_UP_10 (trailing 2 weeks) BBAND_10 (trailing 2 weeks) BBAND_DOWN_10 (trailing 2 weeks)  BBAND_UP_45 (trailing 2 months) BBAND_45 (trailing 2 months) BBAND_DOWN_45 (trailing 2 months)  BBAND_UP_252 (trailing year) BBAND_252 (trailing year) BBAND_DOWN_252 (trailing year)	Envelopes plotted at a standard deviation level above and below a simple moving average of the price. Because the distance of the bands is based on standard deviation, they adjust to volatility swings in the underlying price.
Simple Moving average	SMA_10 (trailing 2 weeks) SMA_45 (trailing 2 months) SMA_252 (trailing year)	Average of a selected range of closing prices.
Exponential moving average	EMA_10 (trailing 2 weeks) EMA_45 (trailing 2 months) EMA_252 (trailing year)	Moving average that places a greater weight and significance on the most recent data points. An exponentially weighted moving average reacts more significantly to recent price changes than a simple moving average.
Residual Volatility	residual_volatility	Measure of unsystematic risk (risk specific to underlying security)
Typical price	Typical price	Arithmetic average of the high, low, and closing prices for a given period.

# APPENDIX - Momentum Indicators I

Momentum Indicators	Feature name used in regression model	Definition
Rate of Change	ROC_10 (trailing 2 weeks) ROC_45 (trailing 2 months) ROC_252 (trailing year)	Measures the percentage change in price between the current price and the price a certain number of periods ago. The indicator can be used to spot divergences, overbought and oversold conditions, and centerline crossovers.
Relative Strength Index	RSI_10 (trailing 2 weeks) RSI_45 (trailing 2 months) RSI_252 (trailing year)	Measures the magnitude of recent price changes to evaluate overbought or oversold conditions in the price of a stock or other asset. RSI values of 70 or above indicate that a security is becoming overbought or overvalued and may be primed for a trend reversal or corrective pullback in price. An RSI reading of 30 or below indicates an oversold or undervalued condition.
Stochastic Fast K	STOCHF_K_10 (trailing 2 weeks) STOCHF_K_45 (trailing 2 months) STOCHF_K_252 (trailing year)	The Stochastic Oscillator is a momentum indicator that shows the location of the close relative to the high-low range over a set number of periods. The Stochastic Oscillator is displayed as two lines. The main line is called "%K." Stochastic oscillators tend to vary around some mean price level, since they rely on an asset's price history.
Stochastic Fast D	STOCHF_D_10 (trailing 2 weeks) STOCHF_D_45 (trailing 2 months) STOCHF_D_252 (trailing year)	Moving average of Stochastic Fast K indicator



# APPENDIX - Momentum Indicators II

Momentum Indicators	Feature name used in regression model	Definition
Stochastic K	STOCH_K_10 (trailing 2 weeks) STOCH_K_45 (trailing 2 months) STOCH_K_252 (trailing year)	Some traders find that stochastic fast indicator is too responsive to price changes, which ultimately leads to being taken out of positions prematurely. The slow stochastic is a 15 period moving average to the %K of the fast calculation.
Average Directional Movement Index	ADX_10 (trailing 2 weeks) ADX_45 (trailing 2 months) ADX_252 (trailing year)	The average directional index (ADX) determines the strength of a trend. The trend can be either up or down, and this is shown by two accompanying indicators, the minus directional indicator (-DI) and the plus directional indicator (+DI).
Plus directional indicator	PLUS_DI_10 (trailing 2 weeks) PLUS_DI_45 (trailing 2 months) PLUS_DI_252 (trailing year)	The Plus Directional Indicator is a component of the Average Directional Index (ADX) and is used to measure the presence of an uptrend. When the +DI is sloping upward, it is a signal that the uptrend is getting stronger.
Minus directional indicator	MINUS_DI_10 (trailing 2 weeks) MINUS_DI_45 (trailing 2 months) MINUS_DI_252 (trailing year)	The Negative Directional Indicator measures the presence of a downtrend and is part of the Average Directional Index (ADX). If -DI is sloping upward, it's a sign that the price downtrend is getting stronger.
Chande momentum oscillator	CMO_10 (trailing 2 weeks) CMO_45 (trailing 2 months) CMO_252 (trailing year)	The Chande momentum oscillator uses momentum to identify relative strength or weakness in a market. It calculates the difference between the sum of recent gains and the sum of recent losses and then divides the result by the sum of all price movements over the same period.