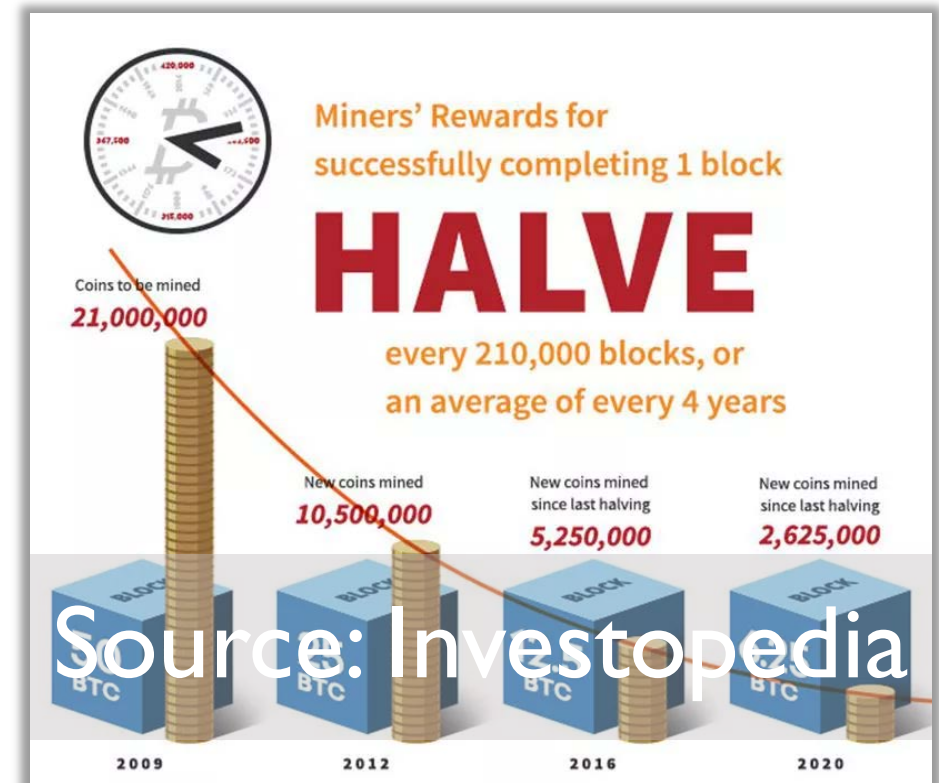# DATA MINING AND PREDICTIVE MODELING FOR CRYPTO-CURRENCIES

LAURA DEANS, JAMES BEEDLE, THEO SHIN, INDERPAL DHILLON
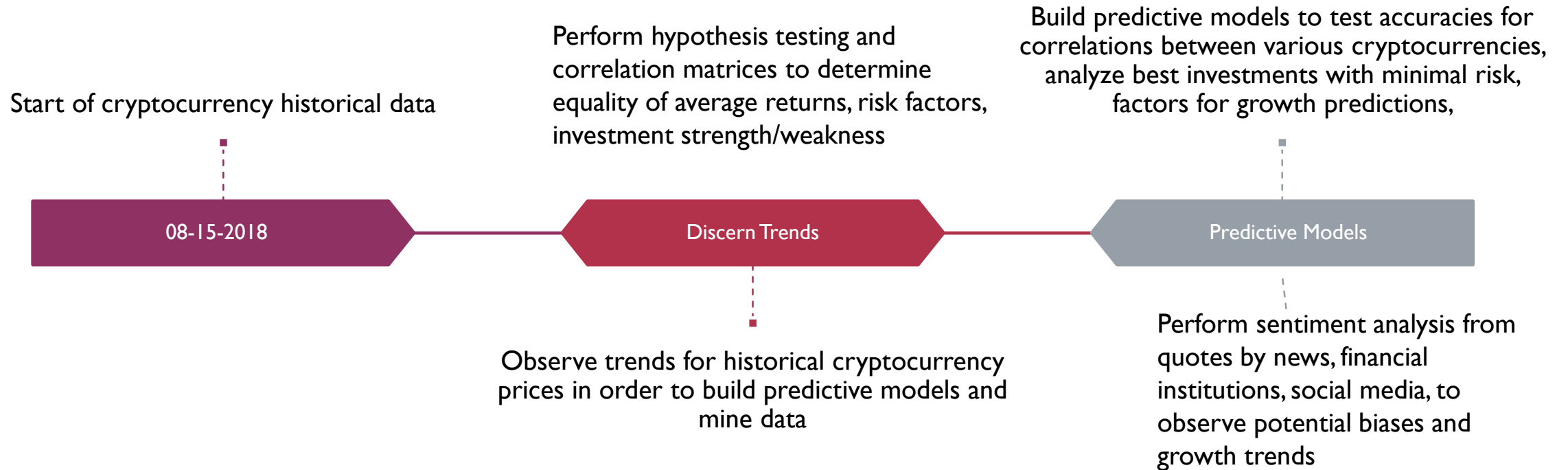
INTRODUCTION: BITCOIN MINING

KEY TAKEAWAYS

- By mining, you can earn cryptocurrency without having to put down money for it.

- Bitcoin is mined in units called "blocks."

- Double spending means, as the name suggests, that a Bitcoin user is illicitly spending the same money twice.

- You need either a GPU (graphics processing unit) miner or an application-specific integrated circuit (ASIC) miner.

- Mining rewards are paid to the miner who discovers a solution to the puzzle first, and the probability that a participant will be the one to discover the solution is equal to the portion of the total mining power on the network.

Miners' Rewards for successfully completing 1 block

HALVE

every 210,000 blocks, or an average of every 4 years

Coins to be mined 21,000,000

New coins mined 10,500,000

New coins mined since last halving 5,250,000

New coins mined since last halving 2,625,000

2009   2012   2016   2020

Source: Investopedia

PROJECT DESCRIPTION

Start of cryptocurrency historical data

Perform hypothesis testing and correlation matrices to determine equality of average returns, risk factors, investment strength/weakness

Build predictive models to test accuracies for correlations between various cryptocurrencies, analyze best investments with minimal risk, factors for growth predictions,

| 08-15-2018 | Discern Trends | Predictive Models |

Observe trends for historical cryptocurrency prices in order to build predictive models and mine data

Perform sentiment analysis from quotes by news, financial institutions, social media, to observe potential biases and growth trends

PROJECT DESCRIPTION (CON'T)

- • As the timeline suggests, we are focused on using historical crypto-currency quotes to discover real-time and future trends by employing predictive models (i.e. ANOVA Hypothesis tests and Naïve Bayes classifier). Some questions and deep meaning we would like to discover include which currency provides the best average returns for investments, are trends consistent amongst various groups/countries, unique correlations which can potentially be linked to predictions, can we predict future trends?

PRIOR WORK

FORECAST MODEL

https://towardsdatascience.com
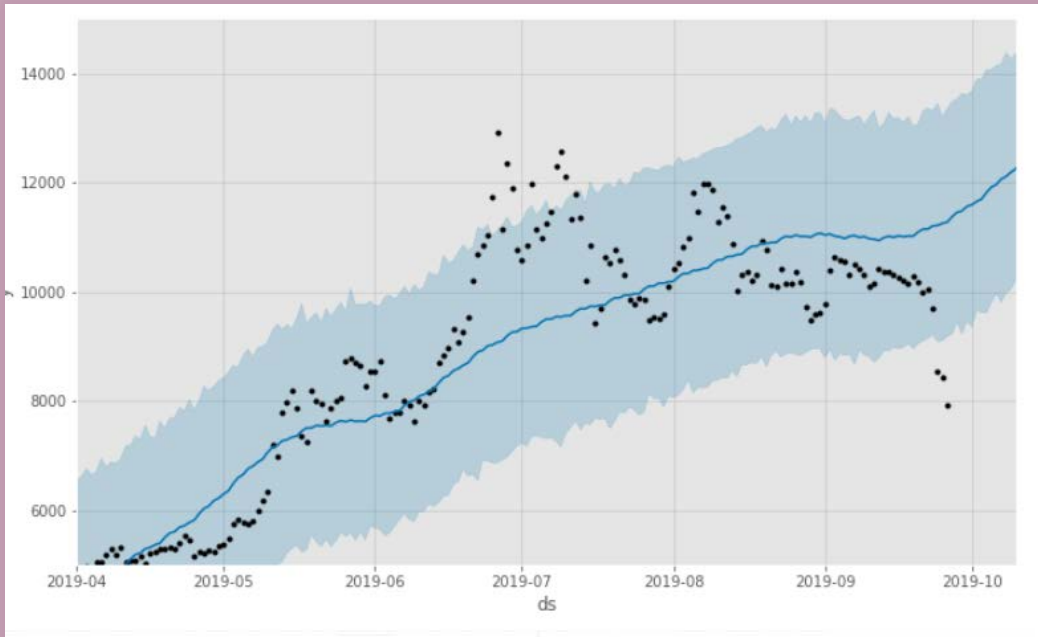
- From Kaggle .csv download @ https://www.kaggle.com/kerneler/starter-ticks-bitcoin-8fbbc257-6. Previous exploratory analysis by combining data and preliminary analysis using a correlation and scatter matrix.

- From https://towardsdatascience.com, forecasted future prices using Time Series models. Author concludes "it is close to impossible to predict the future of Bitcoin, but with **machine learning**, (fit and train models) we can understand where it might go with a high degree of confidence."

DATASETS: CRYPTO-CURRENCY

- Historic Crypto Ticks: Bitcoin, Ethereum, Litecoin, Ripple:

https://www.kaggle.com/albala/ticks-bitcoin-ethereumlitecoin-ripple
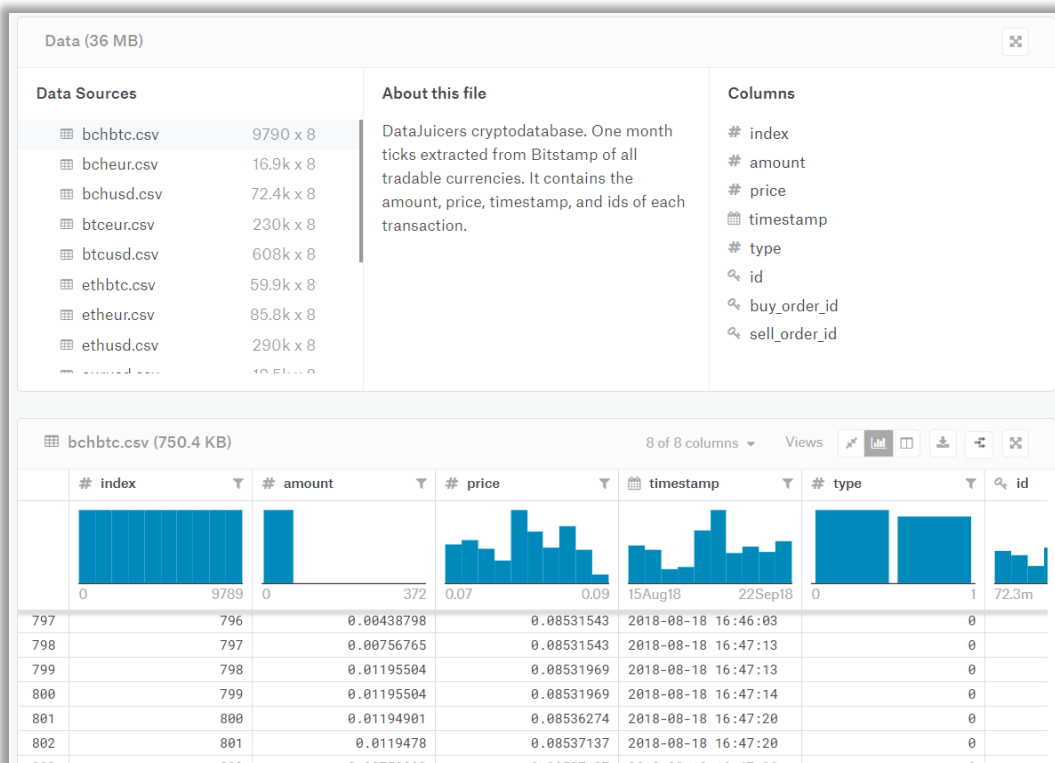
- Cryptocurrency Market Data:

Historical Cryptocurrency Prices For ALL Tokens!

https://www.kaggle.com/jessevent/all-crypto-currencies

- Data Cleaning: We want to aggregate cryptocurrency data for several months at various times. Remove any unnecessary or redundant information along with filling in any missing data using appropriate means.

- Data Pre-Processing: preprocessing using sklearn, reorganize the integration of dual databases to simplify correlation analysis, group cryptocurrencies by most effective trends. Employ visualizations and preliminary analysis to determine potential future growth. A popular visual data visualization normally used with stock market prices is the Candlestick Chart.

- Integrate Data: Combine data from downloaded datasets

PROPOSED WORK: What do we need to do?

LIST OF INTENDED TOOLS

Programs

- Python
- SQL
- Price Converter (Excel)

Tools

- Sklearn
- Seaborn
- nltk
- Wordcloud
- Candlestick
- Matplotlib
- Statistical methods: Naïve Bayes, supervised learning

EVALUATION

- Given the start of the dataset in 2015, we can compare our forecasted predictions to actual results.

- Determine whether hypotheses are accurate based on current trends

- Use forecasting models to observe any correlations between predictions and real-time results

- Compare our prediction models and analysis to relevant data science accuracy measures