* **Introduction**
  + Bitcoin is a digital currency, also known as cryptocurrency, that can be exchanged for goods or services with vendors who accept Bitcoin as payment.
  + It was introduced in 2009 by an anonymous creator known as Satoshi Nakamoto and is the first successful cryptocurrency and uses blockchain technology to secure and verify transactions.
  + The Bitcoin network is public and open-source, which means that anyone can participate.
  + In fact, unlike traditional currencies issued by central banks or governments, Bitcoin is decentralized and transactions take place directly between users through a peer-to-peer network, without the intervention of central authorities.
  + Transactions are authenticated through Bitcoin's proof-of-work consensus mechanism, which rewards cryptocurrency miners for validating transactions.
  + The total number of Bitcoins in circulation is limited to 21 million, which means it cannot be inflated or manipulated in any way and its price is determined by the exchange of supply and demand.
* **Goal**
  + Is it possible to do price forecasting of Bitcoin using machine learning methods in combination with the technical features of its blockchain?
* **Dataset**
  + The dataset containing the Bitcoin information is freely available on the Blockchain.com website
  + The dataset contains ~100k records ❓ describing from 2016 to the present ❓(with half-hour intervals) the most important features of Bitcoin
  + The features have been divided into 4 main categories:
    - Currency statistics: describe its price trend (e.g. market price, number of bitcoins in circulation...)
    - Block details: describe the technical characteristics of its blockchain (e.g. block size, number of transactions...)
    - Mining information: describe the characteristics of the consensus mode “Pow” (e.g. miners revenue, difficulty...)
    - Network activity: describe the actual use of Bitcoin as a method of exchange of value (e.g. number of transactions made, cost per transaction...)
* **Project pipeline**
  + The project is structured like this
    - Data crawling: Bitcoin data retrieval via API call to Blockchain.com
    - Feature engineering: manipulation and visualization of features
    - Features extraction: consider the features more or less related to the price
    - Model train / validation: to train the models and evaluate them by performing hyperparameter tuning
    - Model testing: Test the models and compare the results to answer the initial question
  + Project carried out with Apache Spark (but during feature engineering I converted the Spark dataframe to a Pandas one to make some plots).
* **Feature engineering**
  + Since there are no missing values there was no need to process the data
  + Since there are no categorical features there was no need to encode the data
  + Since ❓ -> no stationarity
    - Se un TS ha un particolare comportamento nel tempo, c'è un'altissima probabilità che lo segua anche in futuro.
  + It was chosen to add the following features regarding the market price
    - Avg price
    - Medie mobili

**Feature selection**

* + All features / most / less correlated to market price (pearson)
  + Dataset split: 95% -> train / validation | 5% -> test
  + Target feature: market price
* **Train / validation of models pipeline**
  + Algoritmi utilizzati
    - Linear regression
    - Generalized linear regression
    - Random forest regressor
    - Gradient boosting tree
  + Evaluation del modello finale
    - RMSE, MSE, MAE, R2, R2adj
  + Simple model:
  + Hyperparameter tuning:
  + Cross validation:
  + Allenamento del modello finale su tutto il train / validation set
* **Model testing** 
  + Prediction test set
  + Evaluation
* **Conclusions**
  + Vantaggi del dataset time series
  + Considerazione finale sulle features utilizzate
  + Considerazioni finali sui risultati dei modelli
  + Risposta alla domanda iniziale
  + Lavori futuri
    - Utilizzare neural networks e compararli con algoritmi di ml normali