TABLE OF CONTENTS

01

STAR SCHEMA

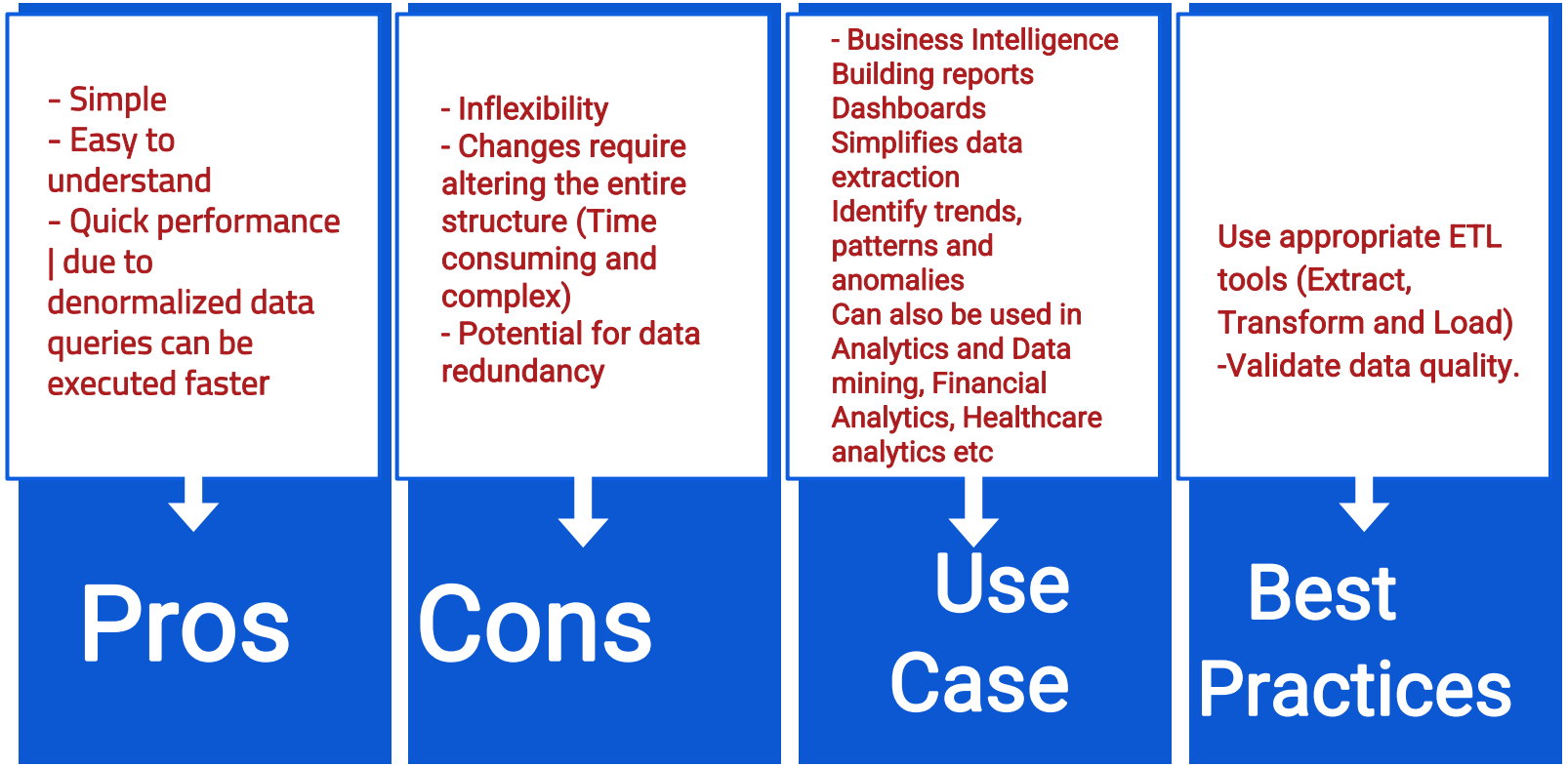
STAR SCHEMA

According to the research of Kimball (2013) the star model has:

A middle table that is the fact table & a set of surrounding tables that are the dimension tables.

The dimension tables are arranged around the main table in a star like arrangement.

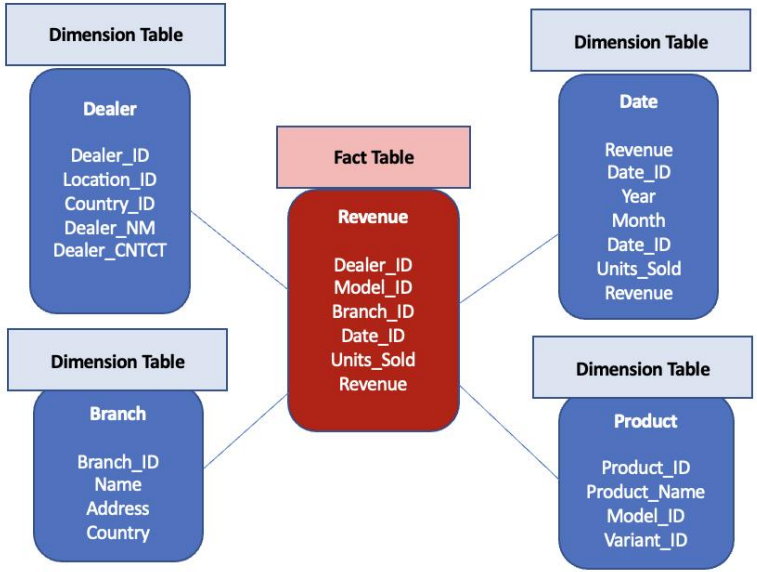
- Fact tables contains numerical values and foreign keys from the surrounding dimension tables.
- Dimension tables contain descriptive information.


- Simple
- Easy to understand
- Quick performance | due to denormalized data queries can be executed faster
Pros
- Inflexibility
- Changes require altering the entire structure (Time consuming and complex)
- Potential for data redundancy
Cons
- Business Intelligence
Building reports
Dashboards
Simplifies data extraction
Identify trends, patterns and anomalies
Can also be used in Analytics and Data mining, Financial Analytics, Healthcare analytics etc
Use Case
Use appropriate ETL tools (Extract, Transform and Load)
-Validate data quality.
Best Practices

02

EXAMPLE

Star Schema Use Case



According to Taylor (2020), the Star Schema can be used for Sales, Financial, Healthcare databases.

The revenue use case has 5 tables. One is a fact tables & the other tables are dimensions tables that describe the fact table.

Revenue table is the Fact Table.

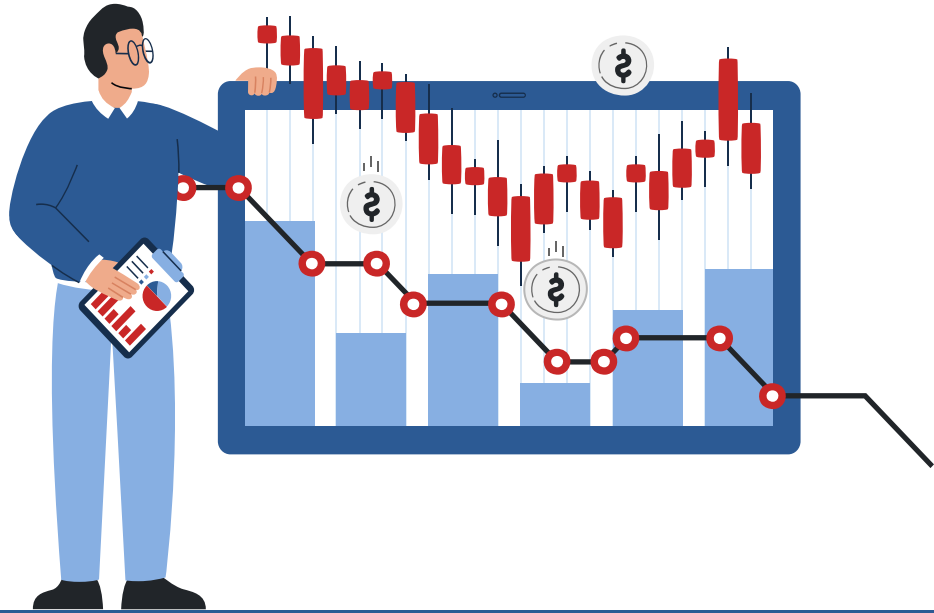
Dealer, Date , Branch and Product tables are the Dimension tables.

Using this datasets we can determine the highest revenue generated by which branch by which dealer in which product and in which time .

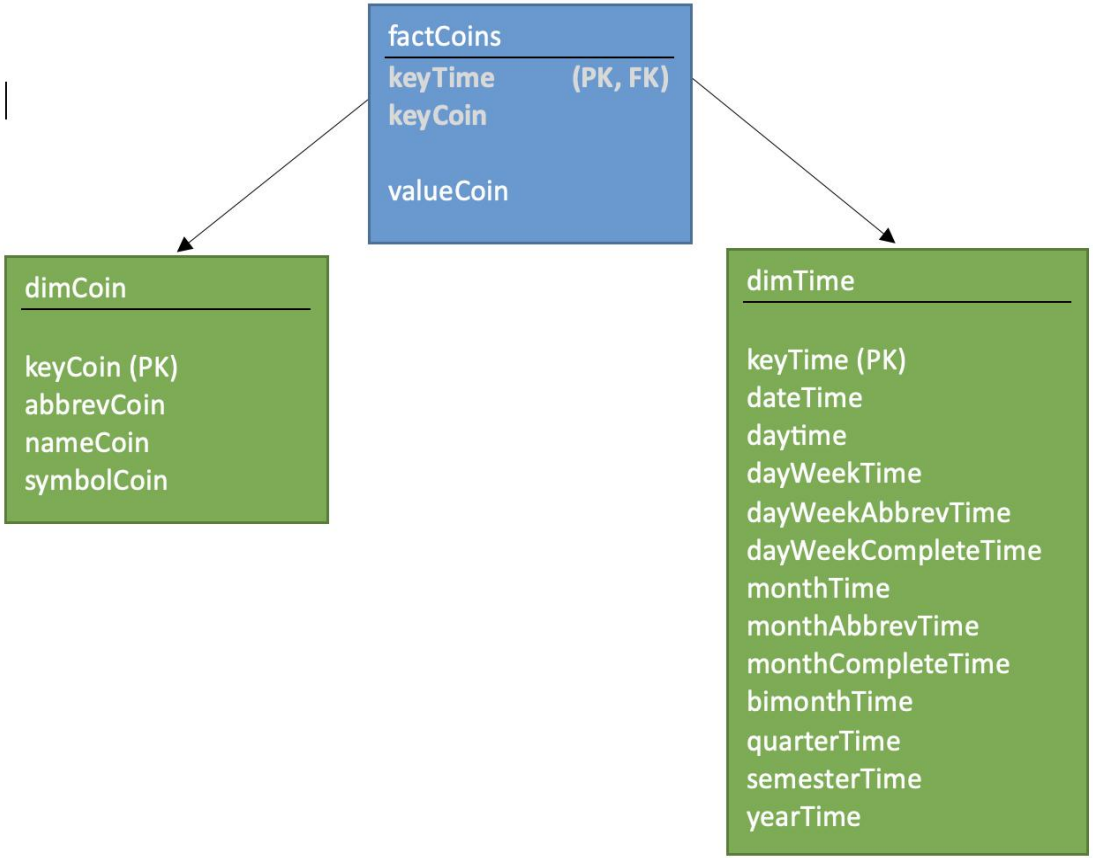
03

DESIGNING THE MODEL

The use case we are using here refers to the dataset we are using from Question 5-9

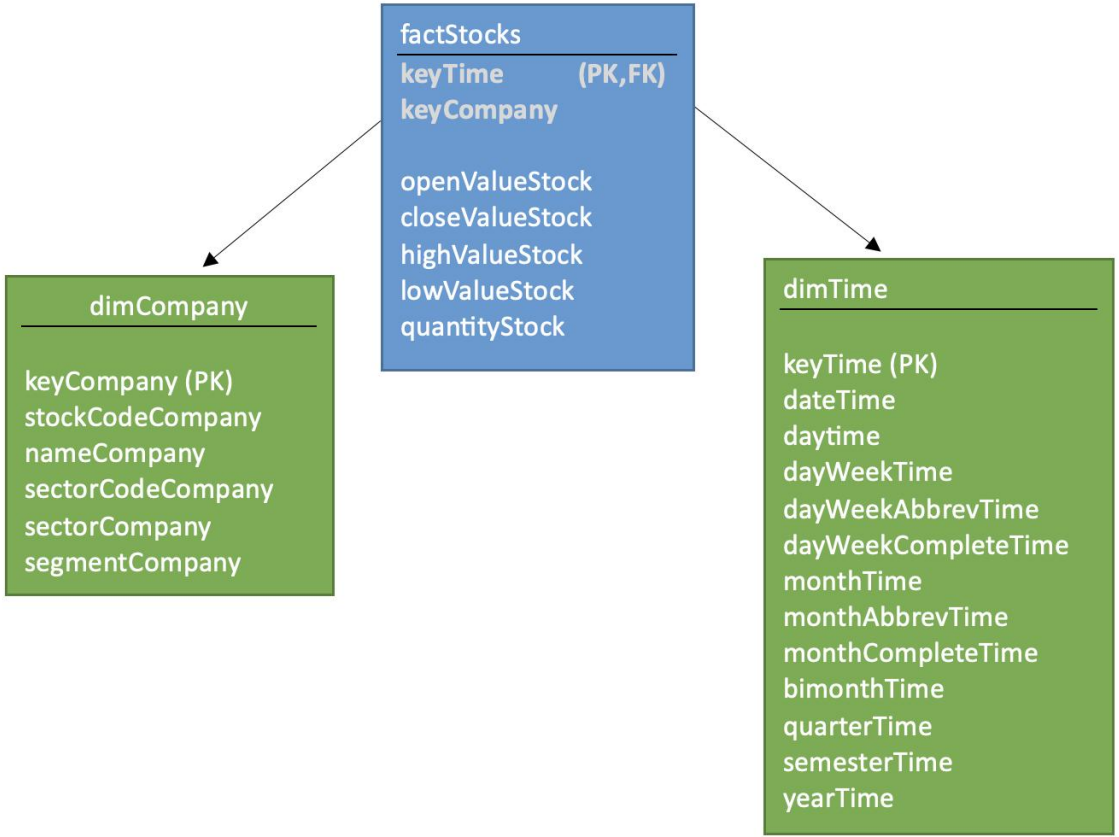


Star Schema 1



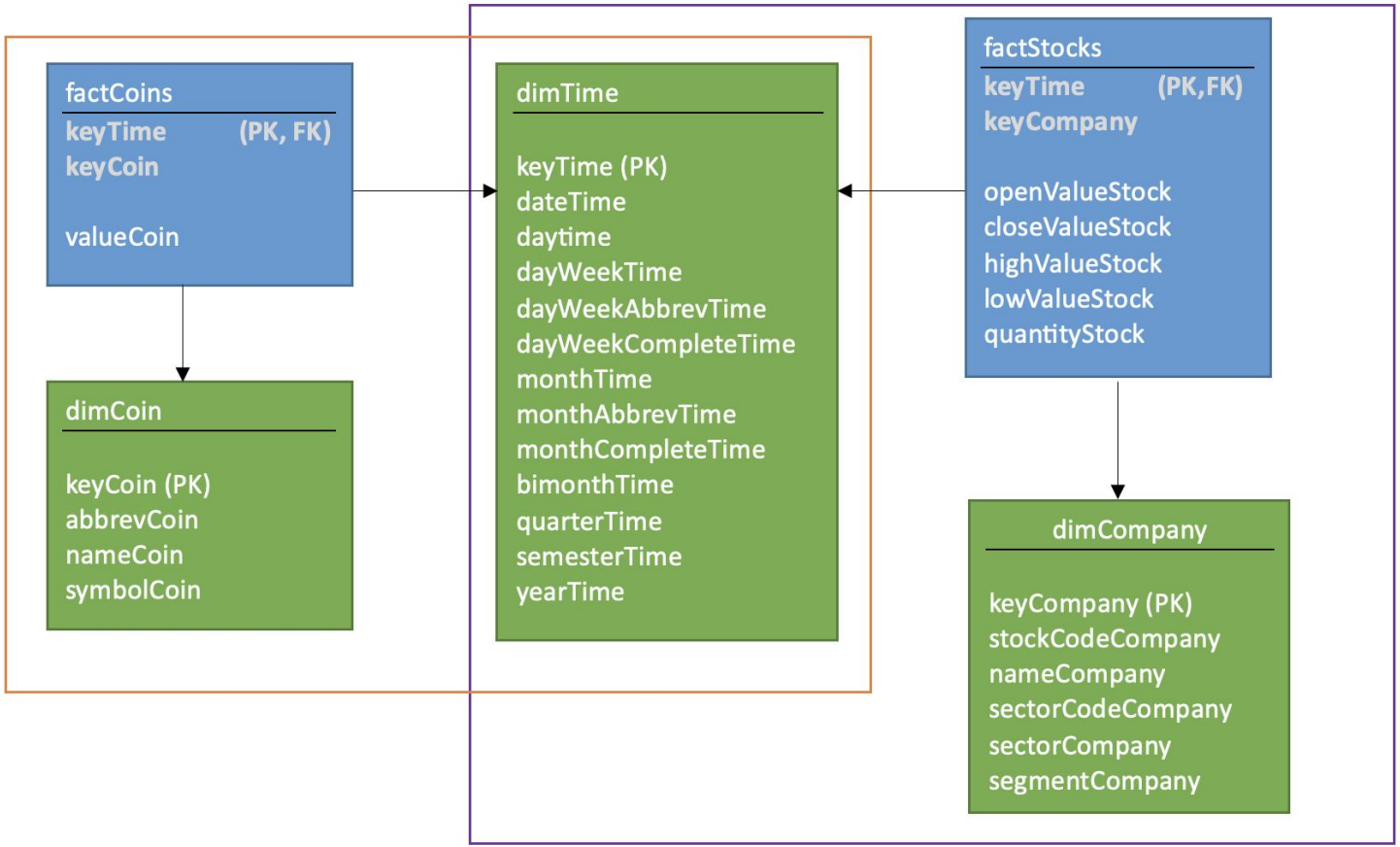
(www.kaggle.com, n.d.)

Star Schema 2



(www.kaggle.com, n.d.)

Star Schema 1+2



(www.kaggle.com, n.d.)

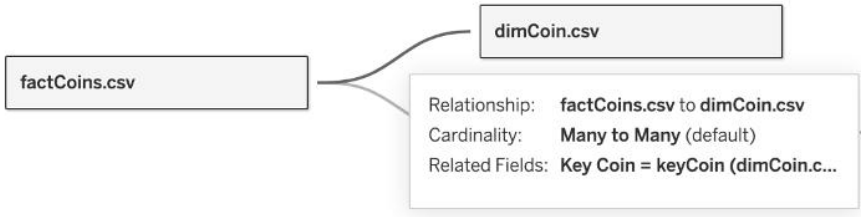
Designing the Star Schema Model for the Brazilian Stock Market in Tableau

How we Designed the Model in Tableau

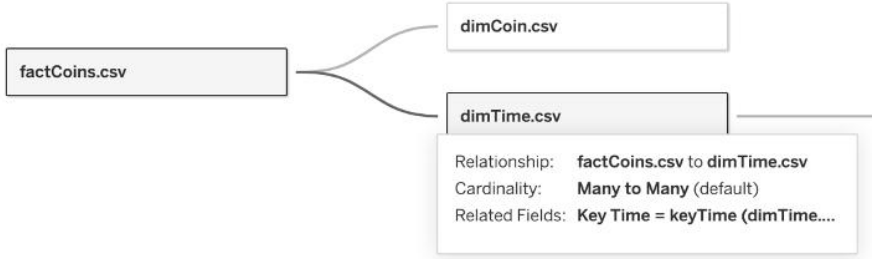
3 easy steps:

1. Connect to the data source
2. Connect factCoins with dimCoin & dimTime
3. Connect factStock with dimCompany & dimTIme

1


factCoins.csv
dimCoin.csv
Relationship: factCoins.csv to dimCoin.csv
Cardinality: Many to Many (default)
Related Fields: Key Coin = keyCoin (dimCoin.c...

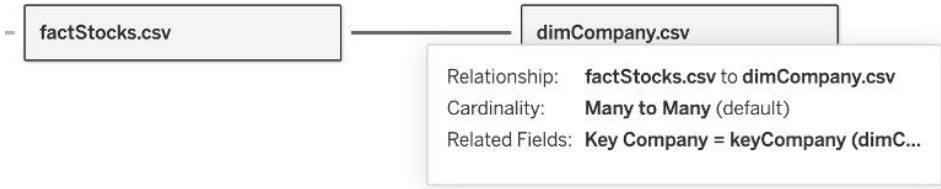
2


factCoins.csv
dimCoin.csv
dimTime.csv
Relationship: factCoins.csv to dimTime.csv
Cardinality: Many to Many (default)
Related Fields: Key Time = keyTime (dimTime....

3


dimTime.csv
factStocks.csv
Relationship: dimTime.csv to factStocks.csv
Cardinality: Many to Many (default)
Related Fields: keyTime (dimTime.... = keyTime (factStock...

4


factStocks.csv
dimCompany.csv
Relationship: factStocks.csv to dimCompany.csv
Cardinality: Many to Many (default)
Related Fields: Key Company = keyCompany (dimC...

Challenges

1. Understanding the stock market basic ideas & terminologies
2. Understanding & finding the relationship between the different tables.
3. Finding the key that will connect the tables together.
4. Huge amounts of data from 1994 – 2020.

04

WHY STAR SCHEMA?

STAR		Historical Data Analysis	High query performance
3NF		Normalization	Higher Processing
DATA VAULT		Real Time Data Requirements	Lack of Correlation Analysis
SNOWFLAKE		Can result in a highly complex database schema for the stock market	Difficult to query across multiple dimensions

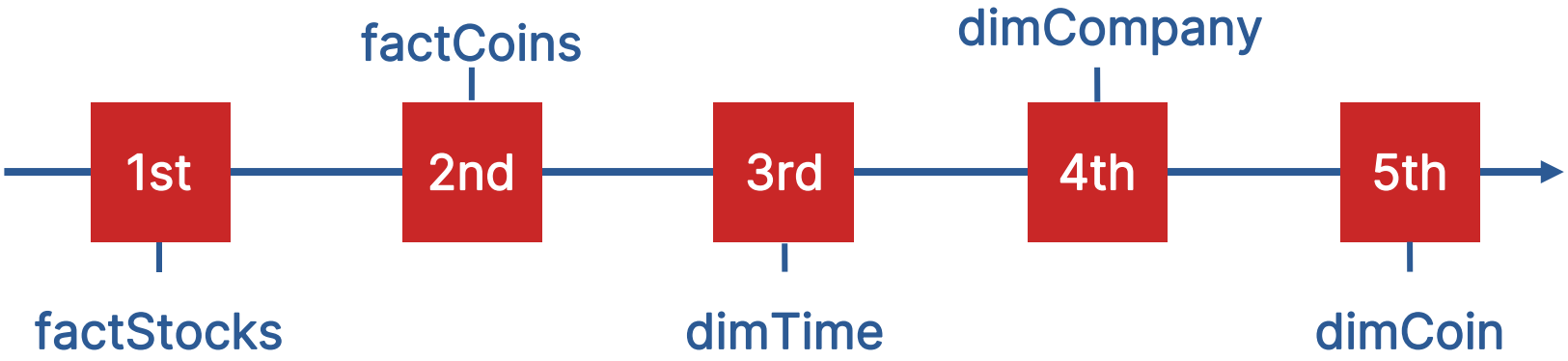
05
Brazilian Stock Market

DATA UNDERSTANDING

About Datasets

Our dataset is based on the Brazilian Stock Market, which offers information on stock prices and numerous organizations from the Brazil Stock Market (1994-2020).

This dataset is split into five tables of which 2 are Fact Tables and 3 are Dimension Tables



(www.kaggle.com, n.d.)

About Variables

factStocks - Fact table with stock prices over time.

keyTime	keyCompany	openValueStock	closeValueStock	highValueStock	lowValueStock	quantityStock
1355	18	45.8	45.8	45.8	45.8	916
1355	107	12.15	12.6	12.6	12.15	668420
1355	108	17	17.5	17.5	16.95	3789172

factCoins- Fact table with coin value over time.

keyTime	keyCoin	valueCoin
4	1	0.94
4	2	2.51
5	1	0.93

dimCoin - Dimension table with information about the coins.

keyCoin	abbrevCoin	nameCoin	symbolCoin
1	USD	DOLAR	$
2	EUR	EURO	,Ç¨

dimCompany - Dimension table with information about the companies.

keyCompany	stockCodeCompany	nameCompany	sectorCodeCompany	sectorCompany	segmentCompany
1	BRAP4	BRADESPAR	IMAT	BM&FBOVESPA BA	SEGMENTS AND SECTORS
2	PMAM3	PARANAPANEMA	IMAT	BM&FBOVESPA BA	SEGMENTS AND SECTORS
3	CSNA3	SID NACIONAL	IMAT	BM&FBOVESPA BA	SEGMENTS AND SECTORS

dimTime - Dimension table with information about the datetime.

keyTime	datetime	dayTime	dayWeekTime	dayWeekAbbrevTime	dayWeekCompleteTime	monthTime	monthAbbrevTime	monthCompleteTime	bimonthTime	quarterTime	semesterTime	yearTime
1	01/07/94	1	6	SAB	SABADO	7	JUL	JULHO	4	3	2	1994
2	02/07/94	2	7	DOM	DOMINGO	7	JUL	JULHO	4	3	2	1994
3	03/07/94	3	1	SEG	SEGUNDA	7	JUL	JULHO	4	3	2	1994

Business Questions

Stock market analysis:
What was the opening, closing, high, low values of different stocks on a particular day?

Currency exchange analysis:
What was the exchange rate of different currencies on a particular day?

Time and date analysis:
Which days of the week had the highest trading volume or exchange rate?

Company and sector analysis:
How did different companies or sectors perform in the stock market?

(www.kaggle.com, n.d.)

06

MODELING

Joining Tables for Analysis

To analyze how different companies or sectors performed in the stock market, we combined the provided datasets using their respective keys.

We combined the tables factStocks, factCoins and dimTime using the 'keyTime' column and with dimCompany using 'keyCompany' (using joins in SQL)

timeKey	datetime	dayTime	dayWeekTime	dayWeekAbbrevTime	dayWeekCompleteTime	monthTime	monthAbbrevTime	monthCompleteTime	bimonthTime	quarterTime	semesterTime	yearTime	companyKey
1355	16/03/98	16	2	TER	TERCA	3	MAR	MARCO	2	1	1	1998	3
1355	16/03/98	16	2	TER	TERCA	3	MAR	MARCO	2	1	1	1998	9
1355	16/03/98	16	2	TER	TERCA	3	MAR	MARCO	2	1	1	1998	11

stockCodeCompany	nameCompany	sectorCodeCompany	sectorCompany	segmentCompany	openValueStock	closeValueStock	highValueStock	lowValueStock	quantityStock	valueCoin
CSNA3	SID NACIONAL	IMAT	BM&FBOVESPA BASIC MATERIALS IN	SEGMENTS AND SECTORS	33.5	35.5	36.7	33.44	78234100	1.92
UNIP6	UNIPAR	IMAT	BM&FBOVESPA BASIC MATERIALS IN	SEGMENTS AND SECTORS	0.29	0.31	0.31	0.28	214100	1.92
GOAU4	GERDAU MET	IMAT	BM&FBOVESPA BASIC MATERIALS IN	SEGMENTS AND SECTORS	42	41.89	42	41.8	171670	1.92

Creation of a New Column for Analysis

Calculated the daily percentage change in stock prices for each company, and included this information as a new column in the new dataset .

Creation of a Pivot Table for Analysis

Aggregated the daily percentage change data by sector to calculate the average performance of each sector over time (using pivot tables)

	A	B
1	**sectorCompany**	**Average of daily percentage change**
2	SPECIAL TAG-ALONG STOCK INDEX (ITAG)	13.02%
3	SPECIAL CORPORATE GOVERNANCE STOCK INDEX (IGC)	22.86%
4	SMALLCAP INDEX (SMLL)	-1.85%
5	NOVO MERCADO CORPORATE GOVERNANCE EQUITY INDEX (IGC-NM)	13.08%
6	MIDLARGECAP INDEX (MLCX)	4.78%
7	BM&FBOVESPA REAL ESTATE INDEX (IMOB)	-6.50%
8	BM&FBOVESPA PUBLIC UTILITIES INDEX (UTIL)	-10.08%
9	BM&FBOVESPA INDUSTRIALS INDEX (INDX)	7.00%
10	BM&FBOVESPA FINANCIALS INDEX (IFNC)	-3.73%
11	BM&FBOVESPA ELECTRIC UTILITIES INDEX (IEE)	1.54%
12	BM&FBOVESPA DIVIDEND INDEX (IDIV)	1.60%
13	BM&FBOVESPA CONSUMER STOCK INDEX (ICON)	2.17%
14	BM&FBOVESPA BASIC MATERIALS INDEX (IMAT)	-0.66%
15	**Grand Total**	**4.57%**

07

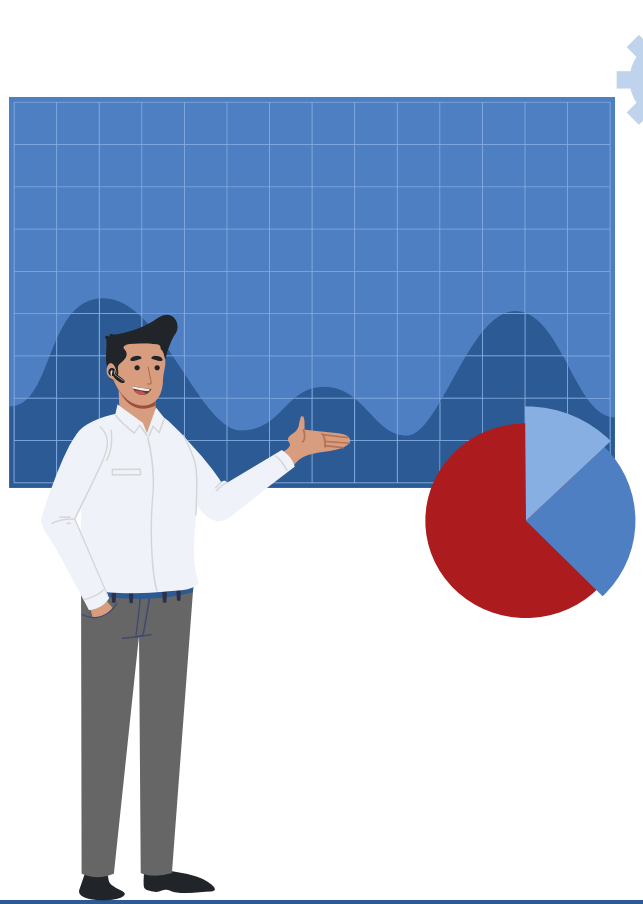
ANALYSIS & EVALUATION

Business Strategies

- Buy & hold (Beers, 2020)
- Value investing (Hayes, 2022)
- Growth investing (Segal, 2019)

Business Requirements

- Transparency
- Accessibility
- Regulation
- Liquidity
- Security



08

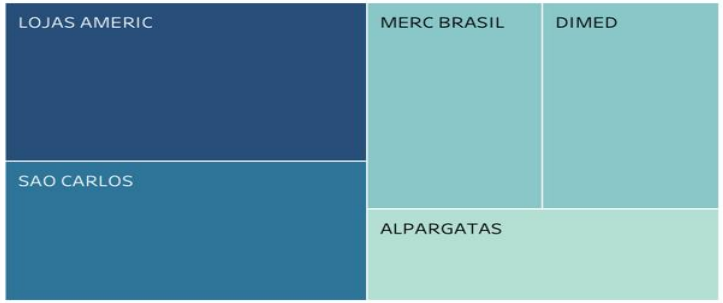
REPORTING

Dashboard representing how different companies or sectors performed in the stock market

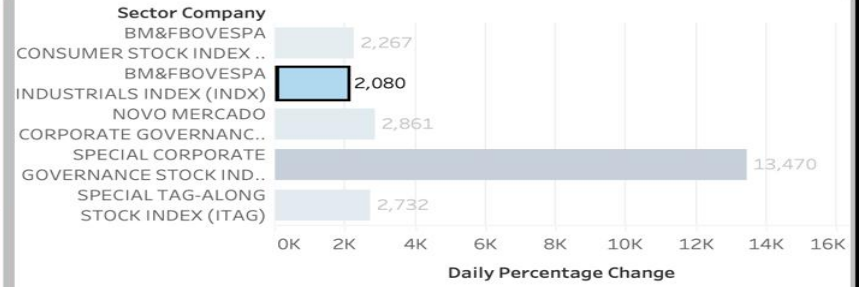
Years

1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

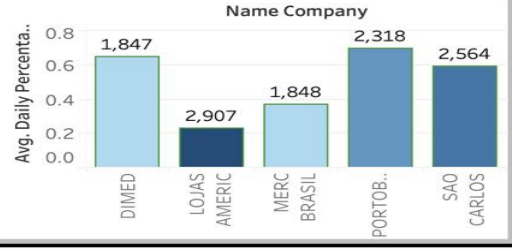
Top companies showing the daily percentage change in stock prices.



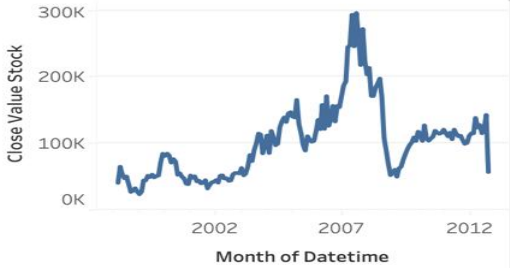
Top sectors companies showing the daily percentage change in stock prices.



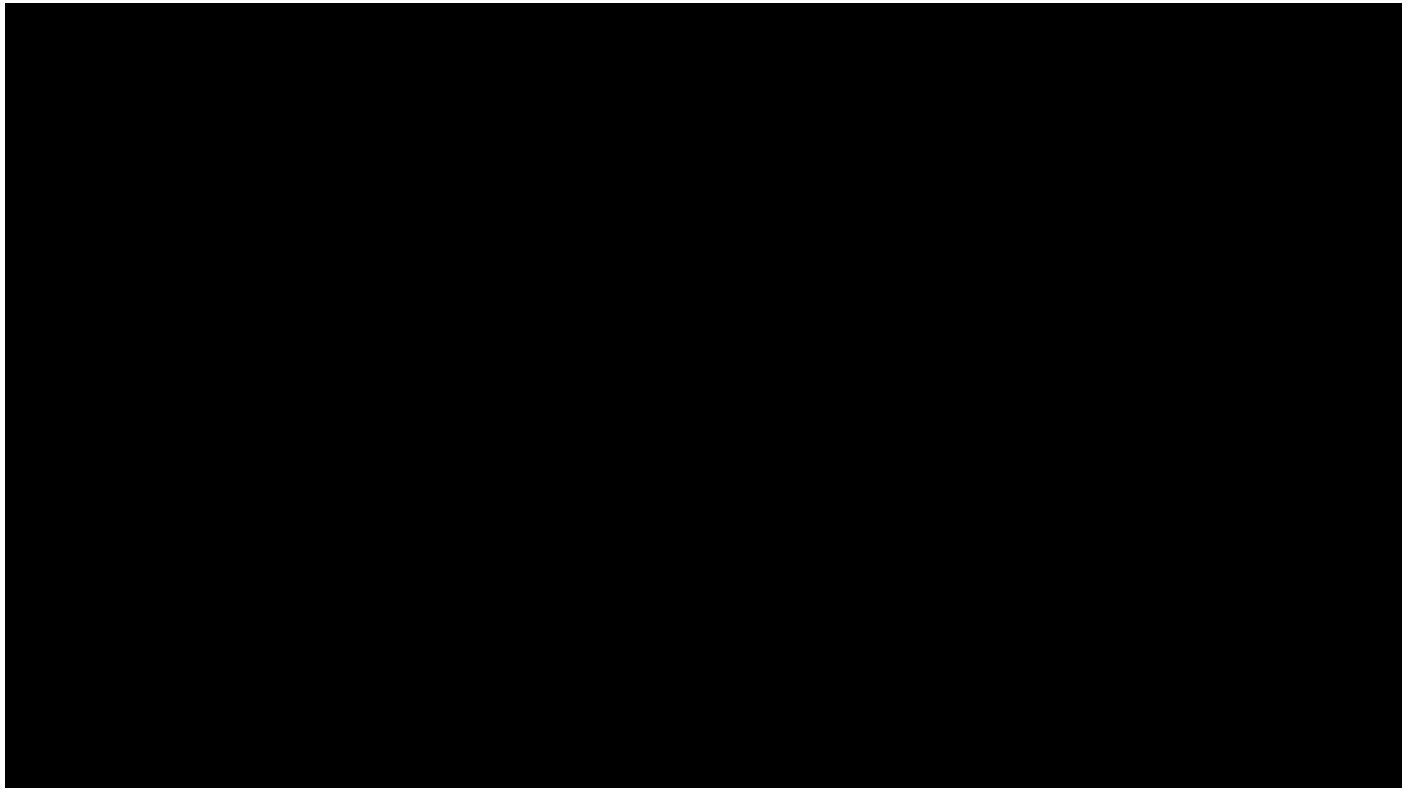
Top Companies showing the average daily percentage change in stock prices.



Trend of close stock value over time



Companies with high stock value



A PICTURE IS WORTH A THOUSAND WORDS

09

RECOMMENDATIONS

Our Recommendations

Based on our analysis and reports, the following overall recommendations can be given:

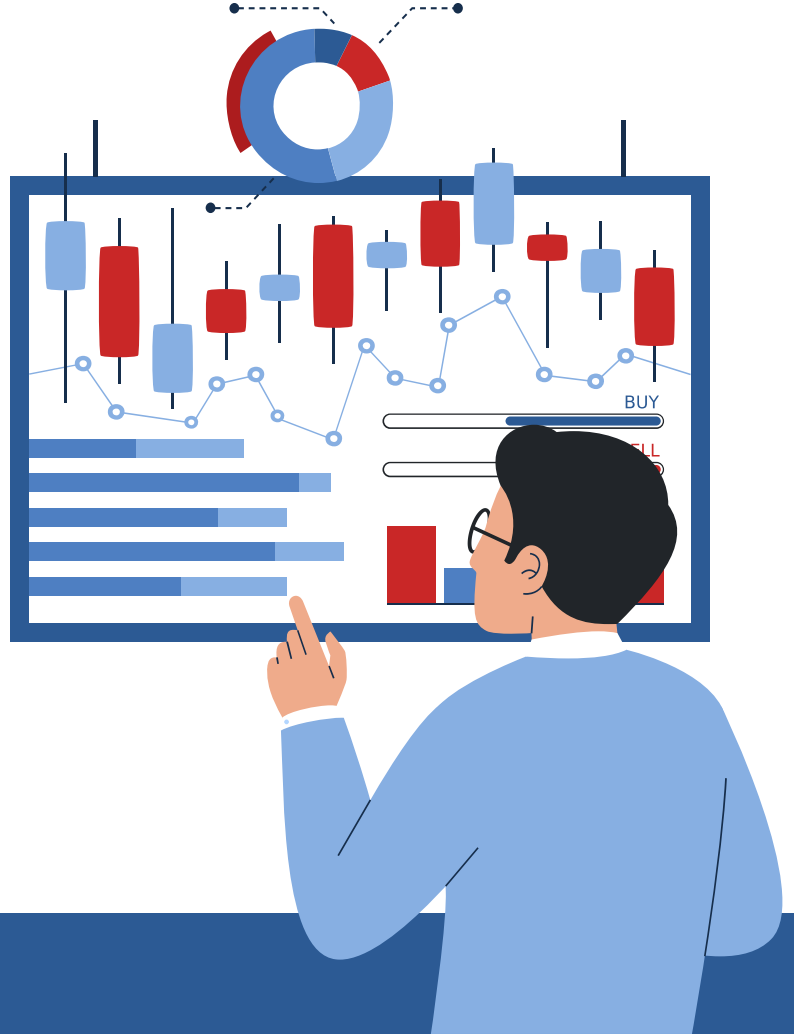
- By diversifying investments across different sectors, investors can mitigate the risk of losses in case of underperformance of a specific sector.

- The daily percentage change in stock prices can indicate the volatility and trend of a particular stock. Investors can consider this information when making investment decisions.

- Investors can consider the correlation between daily percentage change in stock price information when building a portfolio to ensure that their investments are not highly correlated, which can increase risk.

10

CONCLUSION



Conclusion

- Two star models lengthening analysis time.
- Difficulties with data preparation, visualization, and using Tableau.
- Numerous stock market data points.
- A full investigation requires interaction.
- A need for tools for data visualization and programming skills.
- Difficulties in gathering and evaluating datasets are comparable.

REFERENCES

- Beers, B. (2020). *How a Buy-and-Hold Strategy Works*. [online] Investopedia. Available at: https://www.investopedia.com/terms/b/buyandhold.asp.
- Hayes, A. (2022). *Value Investing*. [online] Investopedia. Available at: https://www.investopedia.com/terms/v/valueinvesting.asp.
- Kimball, R. (2013). *The data warehouse toolkit : the definitive guide to dimensional modeling*. Wiley, Sydney.
- Segal, T. (2019). *Is Growth Investing the Right Money-Making Method for You?* [online] Investopedia. Available at: https://www.investopedia.com/terms/g/growthinvesting.asp.
- Taylor, D. (2020). *What is Star Schema in Data Warehouse modeling?* [online] www.guru99.com. Available at: https://www.guru99.com/star-schema-in-data-warehouse-modeling.html.
- Tretina, K. (2021). *How Does The Stock Market Work?* [online] Forbes Advisor. Available at: https://www.forbes.com/advisor/investing/what-is-the-stock-market/.
- www.kaggle.com. (n.d.). *Brazil Stock Market - Data Warehouse*. [online] Available at: https://www.kaggle.com/datasets/leomauro/brazilian-stock-market-data-warehouse [Accessed 8 May 2023].

THANK YOU!