CSI 4142: Phase 1 Deliverable

Conceptual Design – Dimensional Model

Cole Stevens (300171413)
William Beaupre (300174392)
Emiliano Bustamante (300229811)

Student Contributions

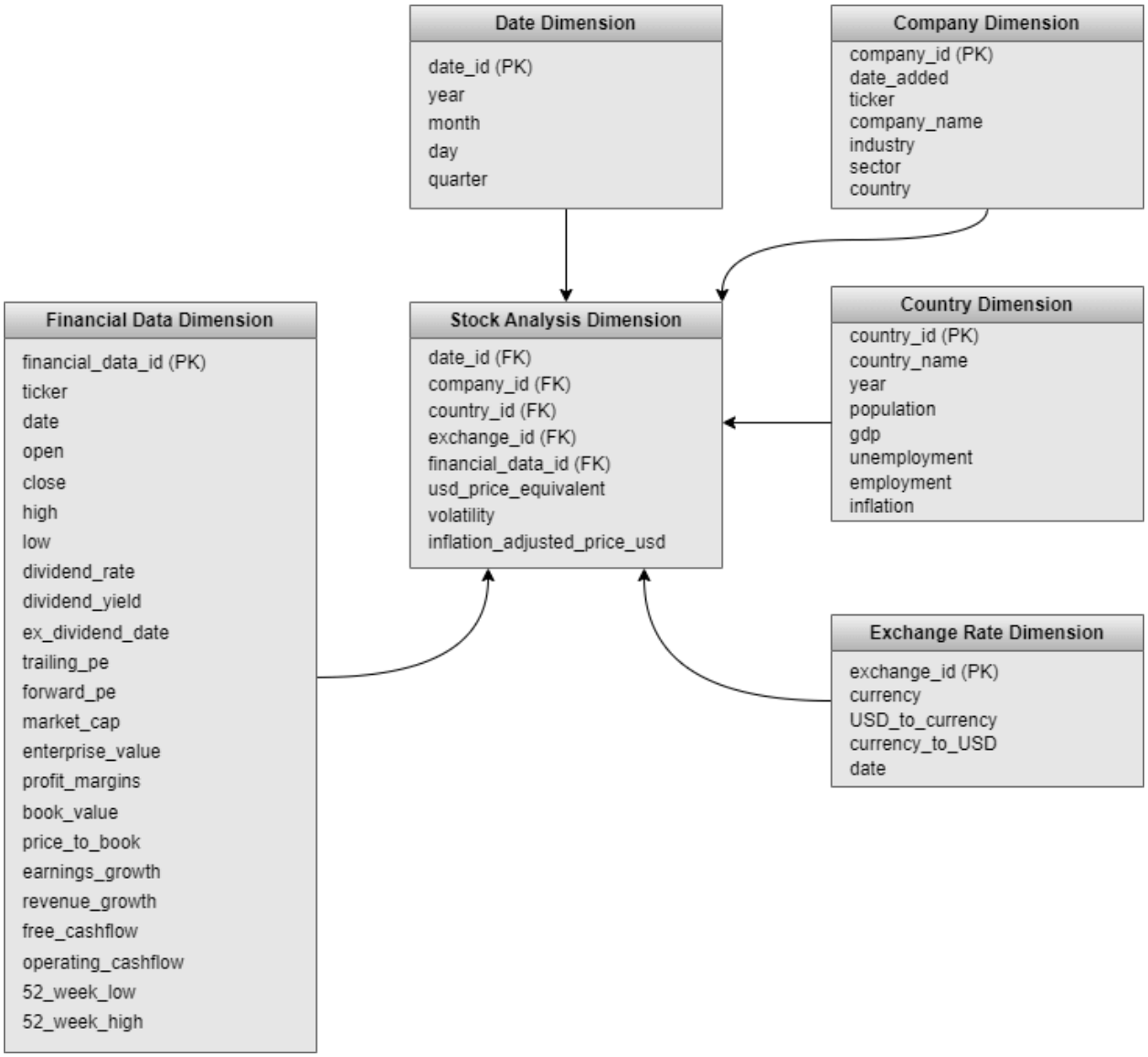
Task/Responsibility	Student(s)
Dimensional Model	Cole Stevens
Grain	William Beaupre
Dimensions & Dimensional Attributes	Cole Stevens
Measures/Facts	William Beaupre
Assumptions	Cole Stevens
Design Mistake Avoidance	Emiliano Bustamante; Cole Stevens

Activity Log

Date	Work Accomplished
2024-02-02	Group initial brainstorm of model dimensions.
2024-02-05	Dimensions sourced and attributes extracted
2024-02-06	Grain determined
2024-02-07	Task distribution finalized; non-sourced dimensions added; assumptions developed
2024-02-08	Design Mistake Avoidance Added

**Note: multiple attempts were made to schedule a meeting with the TA during the week of the delivery, the official plan is still to meet with the TA to discuss our project further before any other deliveries.*

Dimensional Model



Grain

The grain of the project is the volatility of shares, the stock price equivalent in USD, and the modern inflation price equivalent in USD for a given piece of financial data (stock) from a given company in a given country on a given day during specific exchange rates.

Dimensions & Dimensional Attributes

***Note: attributes with an asterisk (*) following them indicate their participation in a primary key.*

Date Dimension

Attribute	Data Type	Minimum	Maximum	Sample
date_id*	integer	19000101	29001231	20240208
year	integer	1900	2900	2024
month	integer	1	12	2
day	integer	1	31	8
quarter	integer	1	4	1

Financial Data Dimension

Attribute	Data Type	Minimum	Maximum	Sample
financial_data_id*	integer	0	unbounded	1233
ticker	String	-	-	'AAPL'
date	String	-	-	'2022-10-12'
open	float	0.0	unbounded	123.65
close	float	0.0	unbounded	175.77
high	float	0.0	unbounded	122.01
low	float	0.0	unbounded	121.87
dividend_rate	float	0.0	1.0	0.96
dividend_yield	float	0.0	1.0	0.0052
ex_dividended_date	String	-	-	'2022-10-12'
trailing_pe	float	0.0	unbounded	29.18818
forward_pe	float	0.0	unbounded	26.175732
market_cap	integer	0	unbounded	2898135613440
enterprise_value	integer	0	unbounded	2904813469696
profit_margins	float	0.0	1.0	0.26163

Financial Data Dimension (cont.)

Attribute	Data Type	Minimum	Maximum	Sample
book_value	float	0.0	unbounded	4.793
price_to_book	float	0.0	unbounded	39.1571
earnings_growth	float	0.0	1.0	0.16
revenue_growth	float	0.0	1.0	0.021
free_cashflow	integer	0	unbounded	86563127296
operating_cashflow	integer	0	unbounded	116433002496
52_week_low	float	0.0	unbounded	143.9
52_week_high	float	0.0	unbounded	199.62
currency	String	-	-	'USD'

Company Dimension

Attribute	Data Type	Minimum	Maximum	Sample
company_id*	integer	0	unbounded	12312
ticker	String	-	-	'AAPL'
company_name	String	-	-	'Apple'
industry	String	-	-	'Consumer Electronics'
sector	String	-	-	Technology
country	String	-	-	United States
date_added	String	-	-	20240205

Country Dimension

Attribute	Data Type	Minimum	Maximum	Sample
country*	String	-	-	United States
year	String	-	-	2011

Country Dimension (cont.)

Attribute	Data Type	Minimum	Maximum	Sample
unemployment	float	0.0	100.0	8.95
employment	float	0.0	100.0	57.326
population	integer	0	unbounded	311583481
gdp	integer	0	unbounded	15599728123000
inflation	float	0.0	unbounded	103.156841568622

Exchange Rate Dimension

Attribute	Data Type	Minimum	Maximum	Sample
exchange_id	integer	0	unbounded	2324
currency	String	-	-	'GBP'
USD_to_currency	float	0.0	unbounded	0.923
currency_to_USD	float	0.0	unbounded	1.077

Measures & Facts

Attribute	Data Type	Minimum	Maximum	Sample
usd_price_equivalent	float	0.0	unbounded	501.05
volatility	float	-100.0	100.0	-0.65
inflation_adjusted_price_usd	float	0.0	unbounded	100.27
date	String	-	-	20240205

Assumptions

- Financial Data from *Yahoo! Finance* (source dataset) produces entries that have the following properties: (1) accurate, (2) recorded within a 24-hour timespan and (3) do not change within that 24-hour timespan (e.g., corrections).
- The countries listed in the Countries dimension do not change names (e.g. Canada does not suddenly become 'Narnia').
- All numeric values included in the Countries dimension are accurate
- All numeric values included in the Companies dimension are accurate, and the String values in the Compantis dimension are categorized properly (e.g., two consumer electronics companies will have the industry column filled out as 'Consumer Electronics', not one as 'Consumer Electronics' and the other being 'Electronics for Consumers').
- Exchange rates only update once per day

Design Mistake Avoidance

Place text attributes in the Fact table.

Certain dimensions (i.e., Financial Data Dimension, Date Dimension, Company Dimension, Country Dimension), could have a text-based attribute serve as the primary key. To avoid this mistake, numeric id attributes were added to these dimension tables to serve as the primary keys, so the composite key for the Fact table (as well as the measures) are numeric.

Ignore the need to track changes

Frequently-changing dimensions (e.g., Financial Data) as well as infrequently-changing dimensions (e.g., Company) have been provided with a date dimension and date_added dimensions respectively to track any changes. In the *Assumptions* section, certain dimensions are assumed to not change at all (e.g., Country), so no date_added dimension is added to prevent unnecessary inflation of data size.

Use operational keys as the primary keys

As mentioned earlier, unique numeric id attributes were assigned to each dimension to prevent any keys being directly sourced from operational databases.

Neglect to declare (and comply with) the grain

Once the grain was declared, all attributes were confirmed to conform to the specified level of granularity.

Neglect a detailed design

The design contains detailed information where applicable to conform to the standards laid out by the project guidelines. Numerous attributes have been included to support in-depth analytical processes, and multiple dimension tables were included to offer a wide range of information types.

References

- World Bank. (2023). *World Bank Open Data.* World Bank. https://data.worldbank.org/
- Yahoo Finance – Stock Market Live, Quotes, Business & Finance News. Yahoo! Finance. https://finance.yahoo.com/
- Bland, G. (2020, June 16). *yfinance Library – A Complete Guide – AlgoTrading101 Blog*. https://algotrading101.com/learn/yfinance-guide
- CPA, S. S., CFA. (2016, March 6). *How to Calculate Volatility of a Stock.* The Motley Fool. https://www.fool.com/investing/how-to-invest/stocks-how-t-calculate-stock-volatility/