HDiSBI - dokumentacja projektu

Kamień milowy 2

Analiza wsteczna zawsze skuteczna Grzegorz Zakrzewski, 313555

11 czerwca 2023

Spis treści

1	Cel projektu	2
2	Zmiany względem KM1	2
3	Finalna architektura rozwiązania	2
4	Opis wykorzystywanych zbiorów danych 4.1 Opis ręcznie przygotowanych plików do zasilenia wymiarów 4.1.1 Wymiar: spółka 4.1.2 Wymiar: waluta 4.2 Opis zewnętrznych źródeł danych do regularnego zasilania tabel faktowych 4.2.1 Yahoo Finance API - dzienne ceny akcji spółek 4.2.2 Bloomberg API - wskaźniki finansowe 4.2.3 NBP API - kursy wymiany walut	3 3 3 4 4 4 5
5	Opis procesu ETL 5.1 Proces ETL - ładowanie wymiarów 5.2 Proces ETL - ładowanie faktów 5.2.1 Fakt: kurs wymiany waluty 5.2.2 Fakt: ceny akcji spółki 5.2.3 Fakt: odczyt wskaźników finansowych spółki	6 6 7 7 8 9
6	Model fizyczny hurtowni danych	10
7	7.1 Kluczowe miary	11 11 11 11
8	Opis warstwy raportowej	11
9	9.1 Raport - analiza techniczna	13 13 14
10	Podsumowanie rezultatów projektu	14
11	Testy funkcjonalne 11.1 Prawidłowe wstawianie pobranych danych do hurtowni 11.1.1 Wymiar: waluta 11.1.2 Wymiar: spółka 11.1.3 Fakt: kurs wymiany waluty 11.1.4 Fakt: dzienne ceny akcji spółki 11.1.5 Fakt: odczyt wskaźników finansowych spółki 11.2 Unikalność rekordów w wymiarze spółki 11.3 Wiarygodność danych w tabeli faktów z dziennymi cenami akcji spółek 11.4 Wiarygodna częstotliwość danych w tabelach faktowych	14 14 15 15 16 17 17 18 18

1 Cel projektu

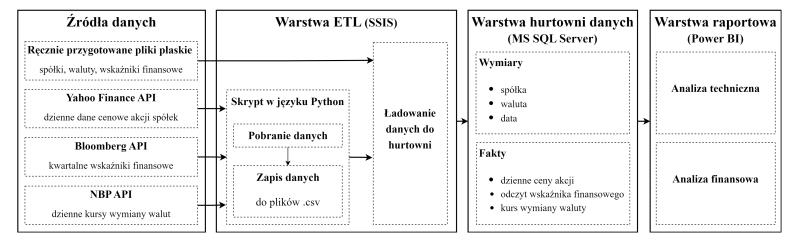
Celem projektu było stworzenie systemu, w którym dane pochodzące z kilku źródeł są przetwarzane, umieszczane w modelu hurtowni danych, a następnie wizualizowane z wykorzystaniem platformy klasy Business Intelligence. Dane obejmują dzienne ceny akcji kilkuset spółek, kwartalne wskaźniki finansowe dotyczące tych spółek, a także kursy wymiany walut. Potencjalnym odbiorcą przygotowanego rozwiązania jest analityk rynków finansowych. System ma pomóc użytkownikowi w analizie sytuacji rynkowej oraz w wydawaniu rekomendacji finansowych.

2 Zmiany względem KM1

Kamień milowy drugi został zrealizowany zgodnie z przyjętymi wcześniej założeniami. Część tej dokumentacji przygotowana w ramach kamienia milowego pierwszego nie uległa znaczącym zmianom. Z tego powodu warto podkreślić, że:

- w opisie wykorzystywanych zbiorów danych zostały zmienione niektóre nazwy kolumn, co ułatwiło przygotowanie procesu ETL;
- opis procesu ETL został doprecyzowany, pojawiły się diagramy procesu ze środowiska SSIS;
- w modelu fizycznym hurtowni kilka kolumn zmieniło typ danych, by lepiej odpowiadać danym wejściowym;
- dodane zostały wymagane sekcje: opis wartswy raportowej, opis realizacji przykładowych raportów, podsumowanie rezultatów projektu, testy funckjonalne.

3 Finalna architektura rozwiązania



Rysunek 1: Diagram architektury rozwiązania

Na rysunku 1 przedstawiona jest finalna architektura rozwiązania. Główne elementy architektury to:

- źródła danych rozwiązanie zasilane jest przez ręcznie przygotowane pliki płaskie oraz dane uzyskane za pomocą różnych API;
- warstwa przetwarzania danych czyli warstwa procesu ETL zaimplementowanego za pomocą narzędzia SQL Server Integration Services; dane dostępne za pomocą API wymagają przygotowania odpowiednich skryptów w języku Python;
- warstwa hurtowni danych wdrożona za pomocą Microsoft SQL Server; zawiera trzy tabele wymiarowe oraz trzy tabele faktowe, uzupełniane za pomocą przetworzonych danych;
- warstwa raportowa przygotowana za pomocą narzędzia Power BI; w jej skład wchodzą dwa raporty.

Każdy z wyżej wymienionych elementów architektury zostanie szczegółowo opisany w odpowiedniej sekcji tej dokumentacji.

4 Opis wykorzystywanych zbiorów danych

Opis wykorzystywanych zbiorów danych jest podzielony na dwie części: opis ręcznie przygotowanych plików służących do zasilenia tabel wymiarowych w hurtowni oraz opis zewnętrznych źródeł danych do regularnego zasilania tabel faktowych.

W tej sekcji kilkukrotnie powtórzą się terminy:

- Yahoo Finance platforma umożliwiająca otwarty dostęp do wiadomości i danych finansowych, dotyczących między innymi spółek giełdowych; na pobieranie danych z tej platformy pozwala między innymi (nieoficjalna) biblioteka yfinance [4], do której będziemy się referować jako Yahoo Finance API;
- Bloomberg płatna platforma dostarczająca dane na temat rynków finansowych, szeroko wykorzystywana
 przez profesjonalnych inwestorów oraz analityków; paczka xbbg [3] umożliwia pobieranie danych z oficjalnego
 API dostarczonego przez ta platformę;
- NBP API udostępniony przez Narodowy Bank Polski serwis [1] do pobierania w formacie JSON kursów wymiany walut.

4.1 Opis ręcznie przygotowanych plików do zasilenia wymiarów

Opisane w tej sekcji pliki zostały przygotowane z myślą o jednorazowym zasileniu wymiarów: spółki i waluty w hurtowni, a także tabeli pomocniczej opisującej wskaźniki finansowe. Oczywiście, ich zawartość nie wyczerpuje w całości wszystkich dostępnych informacji, wręcz przeciwnie, wymiar spółki będzie zawierał kilkaset z kilkudziesięciu tysięcy notowanych przedsiębiorstw, a wymiar waluty tylko sześć spośród ponad setki istniejących jednostek pieniężnych. Przygotowany model hurtowni w żaden sposób nie blokuje potencjalnego dokładania kolejnych rekordów do tych wymiarów.

4.1.1 Wymiar: spółka

Symbol giełdowy, popularnie ticker, to krótki, kilkuznakowy kod służący do jednoznacznej identyfikacji danej spółki akcyjnej na określonej giełdzie [2]. Różne platformy udostępniające dane lub usługi finansowe często korzystają z różnych zestawów tickerów. Na przykład znany producent gier CD Projekt występuje pod tickerem CDR PW Equity na platformie Bloomberg, CDR. WA w serwisie Yahoo Finance, a u wielu brokerów można spotkać się jeszcze z konwencją CDR. PW. Znajomość tickeru jest często niezbędna do pobrania danych finansowych spółki za pomocą API.

	StockSymbol	BloombergTicker	YahooTicker	StockName	StockCurrency	CountryName	IndustryName	SectorName
0	UBSG.CH	UBSG SW Equity	UBSG.SW	UBS Group AG	CHF	Switzerland	Banks Diver	Financial S
1	NESN.CH	NESN SW Equity	NESN.SW	Nestlé S.A.	CHF	Switzerland	Packaged Foods	Consumer De
2	NOVN.CH	NOVN SW Equity	NOVN.SW	Novartis AG	CHF	Switzerland	Drug Manufa	Healthcare
3	ROG.CH	ROG SW Equity	ROG.SW	Roche Holdi	CHF	Switzerland	Drug Manufa	Healthcare
4	KOMB.CZ	KOMB CP Equity	KOMB.PR	Komercní ba	CZK	Czech Republic	Banks Regional	Financial S

Tabela 1: Kilka pierwszych wierszy pliku dim_stock.csv, który posłuży do zasilenia wymiaru spółki

Dysponuję mapą symboli (wykorzystywanych przez jednego brokera) na tickery wykorzystywane w serwisie Bloomberg. Niewielkim wysiłkiem udało się do tej mapy dołożyć tickery wykorzystywane przez Yahoo Finance. Następnie, za pomocą Yahoo Finance API zostały pobrane takie informacje jak nazwa spółki, waluta (w której są notowane akcje), branża, sektor przemysłowy, krótki opis działalności oraz kraj zarejestrowania. W ten sposób udało się stworzyć plik .csv, którego nagłówek przedstawia tabela 1. Ten plik zawiera informacje na temat kilkuset spółek notowanych na różnych giełdach. Te dane posłużyły do zasilenia wymiaru spółki.

Wymiar spółki implementuje mechanizm śledzenia zmian SCD2 ze względu na branżę, sektor przemysłowy oraz opis działajności spółki (IndustryName, SectorName i BusinessSummary). Z tego powodu, do przeprowadzenia testów funkcjonalnych, poza oryginalnym plikiem zawierającym 817 spółek, został przygotowany dodatkowy plik symulujący zmiany w trzech spółkach oraz dodający dwie zupełnie nowe.

4.1.2 Wymiar: waluta

Opisane w poprzedniej sekcji spółki są notowane w sześciu różnych walutach. Ponadto, w przygotowywanym rozwiązaniu przygotowana jest funkcjonalność przeliczania cen akcji i wartości wskaźników finansowanych do jednej,

	CurrencyName	CurrencyCode
0	Swiss franc	CHF
1	Czech koruna	CZK
2	Euro	EUR
3	Polish złoty	PLN
4	Pound sterling	GBP
5	United States dollar	USD

Tabela 2: Dane, które zasilą wymiar waluty

konkretnej waluty "firmowej". Z tego powodu istnieje niewielki wymiar waluty. Zasiliły go ręcznie zebrane dane, zapisane w tabeli 2.

4.2 Opis zewnętrznych źródeł danych do regularnego zasilania tabel faktowych

Opisane w tej sekcji źródła powinny być odpytywane codziennie. W hurtowni są umieszczane dzienne dane cenowe spółek (cena otwarcia, najwyższa i najniższa tego dnia, cena zamknięcie, wolumen) i dzienna częstotliwość odświeżania jest najbardziej naturalna. Warto podkreślić, giełda nie jest czynna w weekendy oraz święta. Podobnie ma się rzecz z kursami wymiany walut. Natomiast czego innego można by się spodziewać po wskaźnikach finansowych. Choć teoretycznie są to dane kwartalne, nie ma jednego konkretnego dnia publikacji wyników finansowych (na podstawie których te wskaźniki są obliczane) dla wszystkich spółek. Dlatego źródło tych danych powinno być również codziennie odpytywane.

4.2.1 Yahoo Finance API - dzienne ceny akcji spółek

Wspomniana wcześniej biblioteka yfinance umożliwia pobranie cen akcji wybranej spółki z określoną częstotliwością, dla wybranego zakresu dat, od razu w formacie ramki danych. Na przykład, wywołanie poniższego kodu:

```
import yfinance as yf
yf.download("CDR.WA", period="max", interval="1d")
```

będzie skutkowało pobraniem następującej ramki danych:

Date	Open	High	Low	Close	Adj Close	Volume
2000-01-03 2000-01-04 2000-01-05 2000-01-06 2000-01-07	36.93 37.59 37.39 37.72 40.52	38.85 38.64 37.97 38.64 43.44	36.76 37.59 37.18 37.72 40.52	38.85 38.01 37.80 37.97	36.65 35.86 35.67 35.83 40.99	46013 90735 54291 67689 67172

Tabela 3: Kilka pierwszych wierszy w przykładowej ramce danych pobranej za pomocą biblioteki yfinance

W zakresie skryptu wykorzystującego Yahoo API leży pobranie danych z od daty ostatniego wpisu w tabeli faktowej dla każdej spółki, usunięcie niepotrzebnej kolumny Adj Close, dodanie kolumny YahooTicker, sformatowanie pola z datą, zmiana nazw kolumn z cenami, a wreszcie połączenie ramek w jeden długi plik .csv. Przykładowe dane po tej transformacji pokazuje poniższa tabela:

	DateKey	OpenPrice_LocalCurrency	HighPrice_LocalCurrency	LowPrice_LocalCurrency	ClosePrice_LocalCurrency	VolumeAmount	YahooTicker
0	20100222	10.6	10.9	10.6	10.8	24444	06N.PL
1	20100223	10.8	10.9	10.6	10.9	9226	O6N.PL
2	20100224	10.7	10.8	10.6	10.8	6178	O6N.PL
3	20100225	10.6	11.3	10.6	11.3	71543	O6N.PL
4	20100226	11.3	11.6	11.2	11.3	23374	O6N.PL

4.2.2 Bloomberg API - wskaźniki finansowe

Biblioteka xbbg pozwala na jednoczesne pobieranie wielu wskaźników finansowych dla listy spółek. Przykład i wynik wywołania prezentuje poniższy kod i tabela:

```
from xbbg import blp
tickers = ['CDR PW Equity']
```

	CDR PW Equity							
	NET_INCOME	CUR_RATIO	EBITDA	OPER_MARGIN	PROF_MARGIN	RETURN_ON_ASSET	SALES_REV_TURN	
2021-12-31	87.518	5.4887	116.718	34.6612	32.0471	8.2752	273.092	
2022-03-31	68.918	5.1482	107.235	39.4738	31.8849	9.5470	216.146	
2022-06-30	44.828	4.4402	73.810	32.2452	27.7109	10.0185	161.770	
2022-09-30	98.704	5.7884	148.845	39.5476	40.2030	13.7007	245.514	
2022-12-31	134.643	5.6428	164.889	43.3753	40.9068	15.6600	329.146	

Tabela 4: Kilka pierwszych wierszy w przykładowej ramce danych pobranej za pomocą biblioteki xbbq

Warto wspomnieć, że wskaźniki pieniężne, to jest $Net\ Income,\ EBITDA$ i Revenue, swoją wartość mają w tej ramce podaną w milionach.

W ramach skryptu ramka danych ulega delikatnym zmianom - dodaniu pola z nazwą tickera, poprawie nazw kolumn. Po takiej transformacji, zapisane do pliku .csv dane będą wyglądają tak:

	${\tt BloombergTicker}$	EBITDA_LocalCurrency	NetIncome_LocalCurrency	${\tt OperatingMargin}$	${\tt ProfitMargin}$	ReturnOnAsset		QuarterDateKey
0	NOVN SW Equity	4193.0	2176.0	21.59	17.12	9.30		20200331
1	NOVN SW Equity	3389.0	1867.0	20.24	16.06	6.01		20200630
2	NOVN SW Equity	3986.0	1935.0	19.24	15.43	5.78		20200930
3	NOVN SW Equity	3651.0	2094.0	20.29	16.07	6.56		20201231
4	NOVN SW Equity	4011.0	2059.0	19.02	16.22	6.50	• • •	20210331

4.2.3 NBP API - kursy wymiany walut

NBP API umożliwia pobranie kursów wymiany walut (względem polskiego złotego) w formacie JSON. Na przykład, skutkiem wejścia w link:

http://api.nbp.pl/api/exchangerates/rates/a/gbp/2002-01-01/2002-01-03/? format=json jest następujący obiekt:

```
}
       "table": "A",
       "country": "W. Brytania",
       "symbol": "789",
       "code": "GBP",
       "rates": [
           {
                "no": "1/A/NBP/2002",
               "effectiveDate": "2002-01-02",
               "mid": 5.7275
10
           },
11
12
                "no": "2/A/NBP/2002",
13
                "effectiveDate": "2002-01-03",
14
               "mid": 5.7091
           }
16
       ]
17
18 }
```

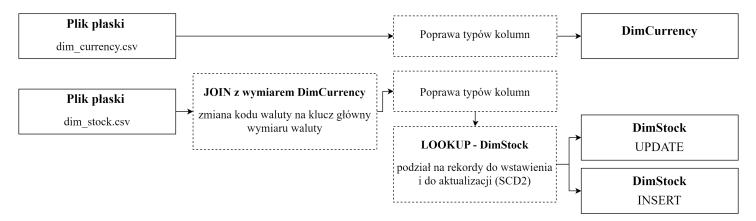
Oczywiście, konieczne jest skonwertowanie pobranego obiektu JSON do innej postaci, na przykład do pliku .csv, którego strukturę przedstawia poniższa tabela:

	ConversionDateKey	SourceCurrencyToPLN_ExchangeRate	SourceCurrency
0	20020102	2.3915	CHF
1	20020103	2.4034	CHF
2	20020104	2.3895	CHF
3	20020105	2.3895	CHF
4	20020106	2.3895	CHF

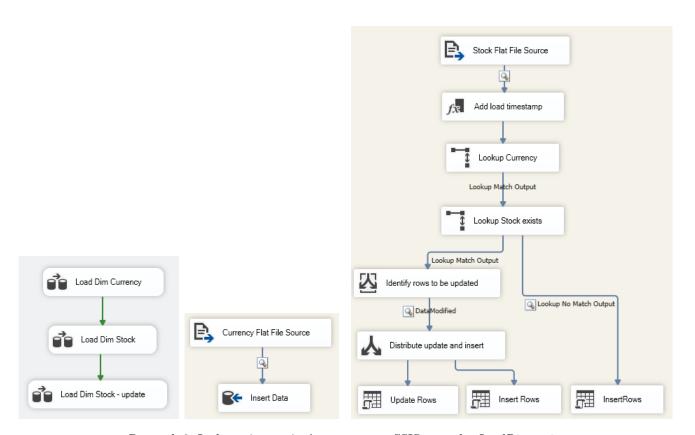
NBP API udostępnia dane od 2 stycznia 2002r., dlatego wszystkie rekordy w tabelach faktowych zaczynają się właśnie od 02.01.2002 r., nawet jeśli dane z Yahoo API lub z Bloomberg API obejmują szerszy zakres dat.

5 Opis procesu ETL

5.1 Proces ETL - ładowanie wymiarów



Rysunek 2: Poglądowy proces ETL dla wymiarów



Rysunek 3: Ładowanie wymiarów za pomocą SSIS - paczka *LoadDimensions* diagram całościowy (lewo), wymiar waluty (środek), wymiar spółki (prawo)

Proces ETL dla wymiarów (diagramy 2 i 3) nie jest skomplikowany - ładowane są jedynie wcześniej przygotowane pliki płaskie z odpowiednimi danymi. Konieczna poprawa typów kolumn uwzględniona jest w kontrolce "Flat File

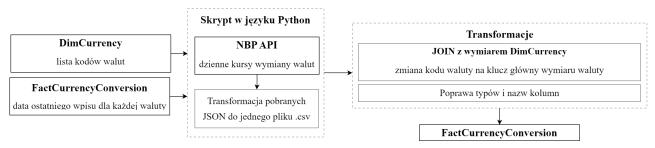
Source". Wymiar spółki wymaga połączenia z wymiarem waluty oraz, z racji mechanizmu SCD2, dodania znacznika czasu oraz podziału na wiersze do wstawienia i do aktualizacji.

5.2 Proces ETL - ładowanie faktów

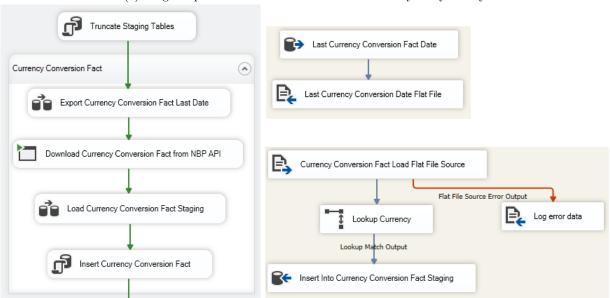
Proces ETL odpowiadający za ładowanie faktów zdefiniowany jest w SSIS w dedykowanej paczce *LoadFacts*. Ten proces dekomponuje się na trzy części, odpowiadające trzem tabelom faktowym.

Ładowanie danych do kazdej tabeli faktowej uwzględnia krok ze skryptem, pobierającym dane z odpowiedniego API i zapisującem wyniki do pliku .csv. Warto podkreślić przebiegającą pod spodem integrację danych. Przebiega ona głównie dzięki wymiarowi spółki, który zawiera tickery z serwisów Yahoo, Bloomberg, a także klucz do wymiaru waluty. Dzięki temu dane pochodzące z trzech serwisów, o różnych zestawach identyfikatorów, mogą stanowić spójną całość w modelu hurtowni danych.

5.2.1 Fakt: kurs wymiany waluty



(a) Diagram procesu ETL dla tabeli faktów - kurs wymiany waluty



(b) Realizacja procesu ETL w SSIS; po lewej - fragment paczki *LoadFacts*; po prawej - eksport daty ostatniego faktu do pliku (góra), ładowanie pobranych danych do tabeli stagingowej (dół)

Na powyższym diagramie zobrazowany jest proces ETL dla tabli faktów przechowującej dane o kursach wymiany walut. Na początku, do pliku .csv eksportowane są daty ostatniego rekordu w tabeli dla każdej waluty. Następuje to z pomocą polecenia SQL:

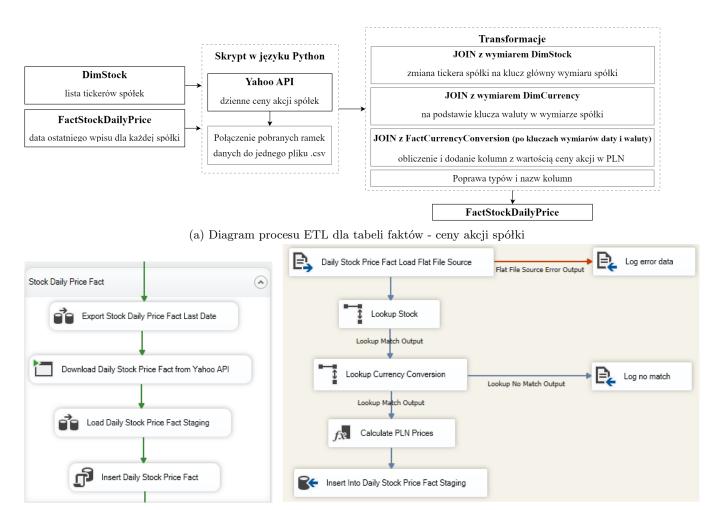
```
1 SELECT
2   dc.CurrencyCode,
3   ISNULL(MAX(dd.Date), CAST('2002-01-01' AS DATE)) AS Date
4 FROM [AWZS].[dbo].[DimCurrency] dc
5   LEFT JOIN [AWZS].[dbo].[FactCurrencyConversion] fcc ON dc.CurrencyKey = fcc.SourceCurrencyKey
6   LEFT JOIN [AWZS].[dbo].[DimDate] dd ON dd.DateKey = fcc.ConversionDateKey
7 WHERE CurrencyCode <> 'PLN'
```

8 GROUP BY dc.CurrencyCode

Wynik tego polecenia trafia do pliku last_currency_conversion_fact_date.csv w określonej lokalizacji. Uruchomiony dalej skrypt (w języku Python) pobiera dane na temat kursów wymiany walut, od dnia odczytanego z pliku. Skrypt zapisuje pobrane dane do pliku currency_conversion_fact_new_load.csv. Następnie zdefiniowane zadanie "Data Flow" wczytuje te dane z pliku, zmienia typy kolumn ("Flat File Source"), pozyskuje klucz główny wymiaru waluty oraz ładuje gotowe rekordy do tabeli stagingowej FactCurrencyConversionStaging. Ostatni krok procesu do załadowanie danych z tabeli stagingowej to tabeli faktowej.

Ładowanie faktów do tabel z dziennymi cenami akcji spółek i z odczytami wskaźników finansowych przebiega na bardzo podobnej zasadzie.

5.2.2 Fakt: ceny akcji spółki



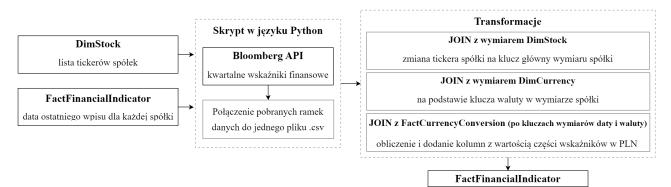
(b) Realizacja procesu ETL w SSIS; po lewej - fragment paczki *LoadFacts*; po prawej - ładowanie danych pobranych z API do tabeli stagingowej

Powyższy diagram pokazuje proces ETL dla tabli faktów przechowującej dane o cenach akcji spółek. Bardzo podobnie jak w przypadku kursów wymiany walut, tutaj tak samo eksportowane są daty ostatnich wpisów dla każdej spółki, skrypt pobiera dane z Yahoo API, następuje ładowanie do tabeli stagingowej poprzedzone transformacjami, a na końcu jest wstawienie do odpowiedniej tabeli faktowej.

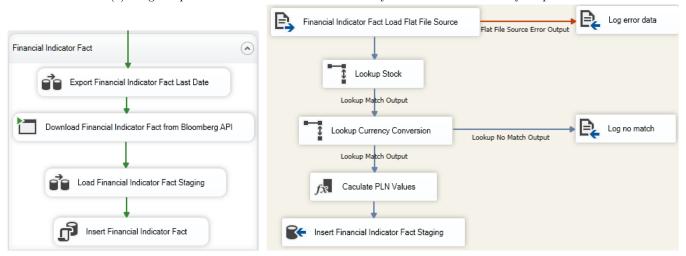
Warto podkreślić jeden krok transformacji. Z użyciem klucza wymiaru waluty (otrzymanego z wymiaru spółki) i klucza wymiaru daty następuje odpytanie tabeli z kursami wymiany walut. Otrzymane wartości służą do obliczenia wartości cen akcji spółki w polskim złotym. W zależności od godziny uruchomienia procesu, może się zdarzyć tak, że ceny akcji spółki (za dzień uruchomienia) są już dostępne przez API, natomiast brak jeszcze kursu wymiany waluty. Takie rekordy nie trafia ją do hurtowni, tylko do odpowiedniego pliku. Ponieważ rekord nie trafia do hurtowni,

następnego dnia zostaje ponownie pobrany z API, lecz wtedy wymagany kurs wymiany waluty prawdopodobnie znajduje się już w tabeli.

5.2.3 Fakt: odczyt wskaźników finansowych spółki



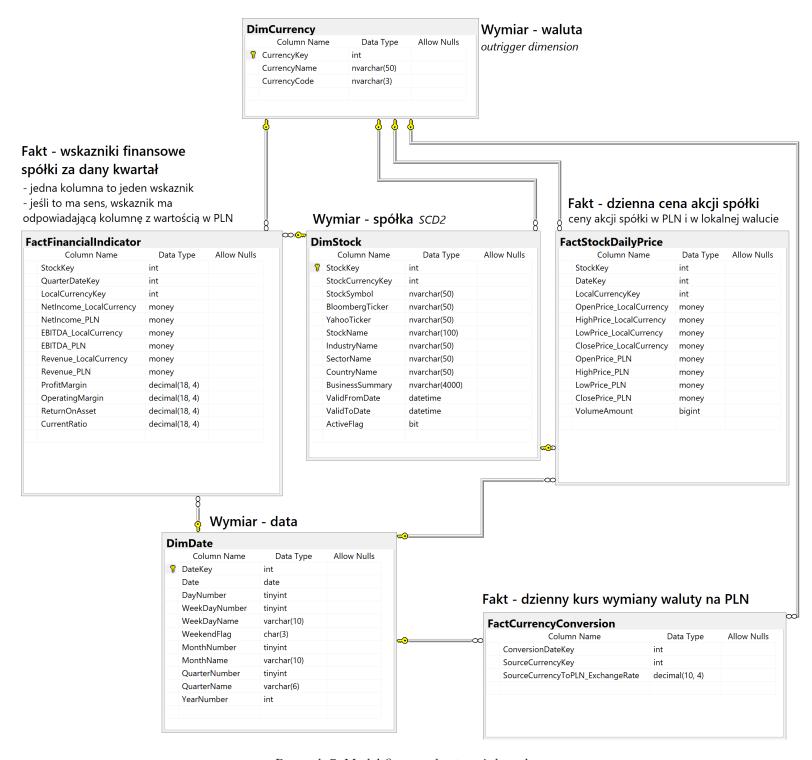
(a) Diagram procesu ETL dla tabeli faktów - odczytu wskaźników finansowych spółki



(b) Realizacja procesu ETL w SSIS; po lewej - fragment paczki *LoadFacts*; po prawej - ładowanie danych pobranych z API do tabeli stagingowej

Proces ETL dla tabli faktów przechowującej dane o odczytach wskaźników finansowych spółek jest bliźniaczo podobny do dwóch poprzednio opisanych procesów, dlatego oszczędzimy kolejny raz tego samego opisu.

6 Model fizyczny hurtowni danych



Rysunek 7: Model fizyczny hurtowni danych

7 Opis kluczowych miar i atrybutów w modelu

7.1 Kluczowe miary

7.1.1 Tabela FactStockDailyPrice - dzienne dane cenowe

OpenPrice_LocalCurrency, HighPrice_LocalCurrency, LowPrice_LocalCurrency, ClosePrice_LocalCurrency to miarki opisujące ceny akcji spółki w ciągu dnia, to jest odpowiednio cenę otwarcia, najwyższą, najniższą oraz cenę zamknięcia. Miara VolumeAmount określa wolumen obrotów (liczba akcji, które zmieniły właściciela). Te miary są kluczowe do przygotowania raportu dotyczącego analizy technicznej.

Opisane miary cenowe podane są w odpowiedniej walucie, w której notowana jest spółka. Te miary mają swój odpowiednik w kolumnach z wartością w polskim złotym, tak aby można było porównywać ceny kilku spółek notowanych w różnych walutach.

Ceny akcji spółki nie są miarami addytywnymi. Jednakże, przy przechodzeniu z częstotliwości dziennej cen na tygodniową ma sens stosowanie takich statystyk jak minimum i maksimum. Wolumen obrotów jest miarką addytywną.

7.1.2 Tabela FactFinancialIndicator - kwartalne dane o wskaźnikach

Kolumny w tej tabeli zawierają wartości wskaźników finansowych. Wskaźniki takie jak Net Income, EBITDA i Revenue mają wartości pieniężne i każdemu wskaźnikowi odpowiadają dwie kolumny: w lokalnej walucie i w polskim złotym. Pozostałe wskaźniki, czyli Profit Margin, Operating Margin, Return On Asset oraz Current Ratio są zdefiniowane jako ciąg operacji arytmetycznych na innych wskaźnikach, reprezentowane są przez liczby rzeczywiste. Odczyty niektórych z tych wskaźników prawdopodobnie można dodawać do siebie, jednak ze względu na planowe wykorzystanie nie będziemy ich traktować jako miary addytywne.

7.1.3 Tabela FactCurrencyConversion - dzienne kursy wymiany walut

Miara SourceCurrencyToPLN_ExchangeRate pełni jedynie rolę pomocniczą. Jest potrzebna do poprawnego działania procesu ETL, który będzie przeliczał waluty lokalne na walutę "firmową", to znaczy polski złoty.

7.2 Kluczowe atrybuty

7.2.1 Tabela DimStock - spółki

- StockName przyjazna dla użytkownika końcowego nazwa spółki;
- BloombergTicker, YahooTicker atrybuty głównie potrzebne do łączenia danych pobieranych w procesie ETL; jednakże, stanowią wartość dodaną dla użytkownika i będą wyświetlane w raporcie;
- IndustryName, SectorName, CountryName oraz BusinessSummary przemysł, sektor, kraj "pochodzenia" oraz podsumowanie biznesowe, czyli atrybuty opisujące spółkę; choć ostatni z tych atrybutów może budzić szczególną wątpliwość, należy pamiętać, że spółek giełdowych jest dosyć sporo; krótkie podsumowanie biznesowe z pewnością ułatwi pracę użytkownikowi.

7.2.2 Tabela DimCurrency - waluty

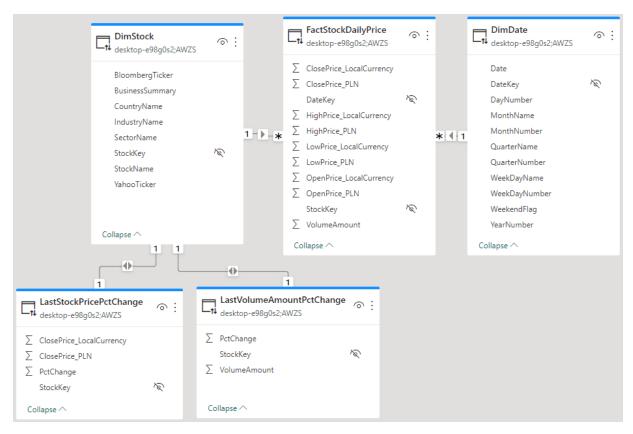
- \bullet CurrencyName pełna nazwa waluty wykorzystanie raportowe;
- CurrencyCode zgodny z ISO 4217 trzyliterowy kod waluty, potrzebny w procesie ETL do pobierania kursów wymiany walut z NBP API.

8 Opis warstwy raportowej

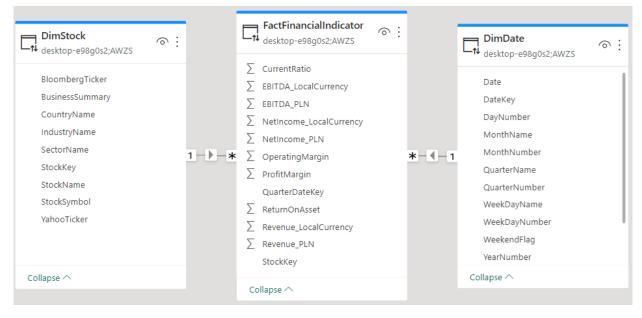
Warstwa raportowa została przygotowana z wykorzystaniem narzędzia Power BI. Efektem pracy są dwa przykładowe raporty opisane w kolejnej sekcji.

Dostęp do danych został zapewniony przez funkcjonalność narzędzia Power BI. Za pomocą opcji "Get data" narzędzie zostało połączone z serwerem SQL. Został wybrany typ połączenia z danymi "DirectQuery", wskutek czego dane nie są przechowywane w narzędziu raportowym.

Model biznesowy danych jest dosyć podobny do modelu fizycznego hurtowni danych. Zostały przygotowane dwa odrębne raporty, które korzystają z różnych tabel w hurtowni danych. Z tego powodu, poniżej zostały umieszczone dwa modele danych. Są to zrzuty ekranu z zakładki "Model view" z narzędzia Power BI.



Rysunek 8: Model danych w pierwszym raporcie



Rysunek 9: Model danych w drugim raporcie

Jak już zostało wspomniane, powyższe modele to podzbiory modelu danych w hurtowni. Usunięte zostały niepotrzebne pola, niektóre kolumny ukryte, oznaczone zostały miarki oraz metody ich agregacji - średnia dla cen akcji,

suma dla wszystkich innych. Jedyna różnice stanowia dwie nowe tabele w modelu 8, które zaraz zostana opisane.

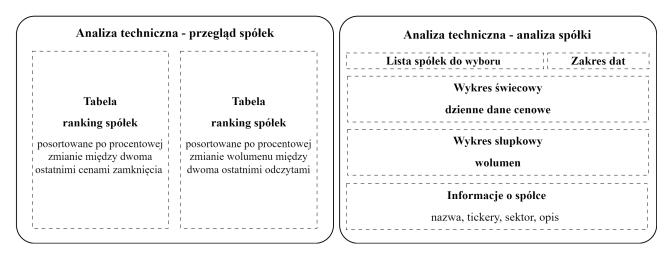
Hierarchie i transformacje nie były kluczowym punktem w implementacji zaplanowanych raportów. Choć można by rozważyć dodanie hierarchii dat lub hierarchii sektor / branża do wymiaru spółki, nie było takiej konieczności. Ceny akcji spółek oraz "pieniężne" wskaźniki finansowe zostały poprawnie oznaczone, aby wyświetlały się jako liczby zmiennoprzecinkowe z określoną precyzją -jeśli można zaliczyć to do transformacji,

Jeden z raportów zakładał przygotowanie rankingu spółek posortowanych po zmianie procentowej między dwiema ostatnimi cenami zamknięcia oraz drugiego, podobnego rankingu, lecz dla wolumenu obrotów. Planowane było dodanie nowej kolumny w warstwie raportowej, przechowującej zmiany procentowe dla tych zmiennych. Taką kolumnę prawdopodobnie można dodać z wykorzystaniem narzędzia Power BI oraz języka DAX, jednakże moje próby takiej implementacji skończyły się fiaskiem. Z tego powodu, w bazie danych zostały dodane dwa widoki: LastStockPricePctChange oraz LastVolumeAmoutPctChange, zbudowane na tabeli faktowej FactStockDailyPrice. Odpytanie tych widoków generuje potrzebne dane. Wynik odpytania tych widoków zawiera klucz wymiaru spółki DimStock, co pozwala na połączenie z resztą modelu danych.

9 Opis realizacji przykładowych raportów

Przygotowane zostały dwa raporty. Przykładowe wydruki raportów są dołączone na końcu tej dokumentacji.

9.1 Raport - analiza techniczna



Rysunek 10: Szkic pierwszego raportu dotyczacego analizy technicznej

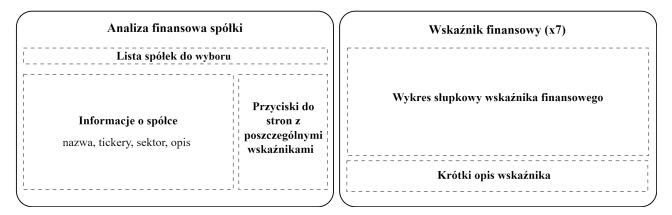
Pierwszy raport, zobrazowany na rysunki 10, ma służyć do analizy technicznej, czyli analizy zmian kursów i wolumenu obrotów akcji spółek. Ten raport ma dwie strony, inaczej mówiąc główne części.

Pierwsza z nich ma za zadanie przedstawić użytkownikowi interesujące na dany dzień spółki. Zrealizowane jest to w formie tabeli, w której są wypisane spółki, posortowane po zmianie procentowej miedzy ceną zamknięcia z dwóch ostatnich dni. Obok znajduje się podobna tabela, lecz w oparciu o zmianę wolumenu obrotów. Spółki można sortować, klikając w nagłówek tabeli. Przygotowanie tej strony w narzędziu Power BI nie było trudne - dzięki dedykowanym widokom (opisanym w poprzedniej części) wystarczyło podpiąć odpowiednie pola.

Druga część raportu opiera się na wyborze jednej, konkretnej spółki, która będzie poddana analizie. Ta część raportu zawiera wykresy: świecowy, przedstawiającym dzienne dane cenowe oraz wykres słupkowy przedstawiający wolumen. Użytkownik ma możliwość wybrać zakres dat, za który chce zobaczyć dane. Ponadto, widoczny jest panel z informacjami na temat spółki (nazwa, tickery, sektor przemysłowy, podsumowanie biznesowe).

Druga strona była trudniejsza w implementacji. Filtry danych - spółka i zakres dat - zostały zrealizowane za pomocą kontrolki "Slicer". Narzędzie Power BI nie umożliwia tworzenia wykresów świecowych, trzeba było dodać ten typ wizualizacji z zewnętrznego źródła - tutaj został wykorzystany "Candlestick by OKViz". Informacje o spółce zostały przedstawione za pomocą pojedynczych kontrolek "Card".

9.2 Raport - analiza finansowa (fundamentalna)



Rysunek 11: Szkic drugiego raportu dotyczącego analizy fundamentalnej

Głównym zadaniem drugiego raportu wyświetlenie historii konkretnego wskaźnika finansowego dla wybranej spółki. Zostało to zrealizowane tak, że główna strona zawiera znane już kontrolki do wyboru spółki oraz przedstawia informacje o wybranym elemencie. Znajduję się też przyciski z nazwami wskaźników finansowych, które przekierowują do odpowiedniej strony w raporcie, bowiem każdy wskaźnik posiada każdą stronę z wykresem słupkowym oraz jednym zdaniem opisu.

W obu raportach planowana była funkcjonalność przełączania wartości pieniężnych pomiędzy polskim złotym a walutą, w której notowana jest spółka. Niestety, nie udało się tego zrealizować. Z tego powodu, pierwszy raport zawiera wartości podane w lokalnej walucie, a drugi - w polskim złotym.

10 Podsumowanie rezultatów projektu

Projekt został zrealizowany zgodnie z założeniami przyjętymi w kamieniu milowym pierwszym. Przygotowana dokumentacja zdaje się pokrywać wszystkie wymagane aspekty. Poniżej tej sekcji znajduje się jeszcze opis przeprowadzonych testów funkcjonalnych oraz przykładowe raporty.

Jak wspomniane było na wstępie, potencjalnym odbiorcą przygotowanego rozwiązania jest analityk rynków finansowych. System ma pomóc użytkownikowi w analizie sytuacji rynkowej oraz w wydawaniu rekomendacji finansowych. Bardziej szczegółowo, wyobrażałem sobie to tak, że analityk zagląda do tego narzędzia codziennie rano i jasno widzi gorące punkty na rynku akcji.

Wykonanie tego projektu mocno zweryfikowało moje wyobrażenia. Przygotowanie narzędzia dla analityka rynków finansowych, zwłaszcza narzędzia tak ogólnego zastosowania, okazało się zadaniem bardzo trudnym. Przygotowane raporty udostępniają zaledwie ułamek możliwości profesjonalnych rozwiązań takich jak Bloomberg Terminal czy Refinitiv Eikon. Zwykła platforma klasy Business Intelligence to zdecydowanie za mało, żeby sprostać oczekiwaniom użytkownika.

Podsumowując, źródła danych, proces ETL, hurtownia danych oraz warstwa raportowa stanowią raczej spójną całość. To gotowe rozwiązanie, które można by nawet wdrożyć, ustawić codzienne odświeżanie i nawet do tego rozwiązania zaglądać. Z punktu widzenia biznesu projekt wypada bardzo słabo i pod względem jakościowym nie umywa się do istniejących narzędzi. Wątpliwe, żeby znalazł się jakiś klient, który zleciłby przygotowanie takiego rozwiązania.

11 Testy funkcjonalne

11.1 Prawidłowe wstawianie pobranych danych do hurtowni

- Co jest testowane? poprawność działania procesu ETL zdefiniowanego w SSIS
- W jaki sposób jest testowane? dla każdej tabeli wymiarów lub faktów uruchamiane są odpowiednie zadania w SSIS, a następnie wykonywane są polecenia SQL w celu weryfikacji zawartości tabeli / ilości wierszy

• Jaki jest oczekiwany wynik? - procesy przechodzą bez problemów (świecą się na zielono), liczba wierszy w hurtowni zgadza się z liczbą wierszy w pliku źródłowym / na diagramie

11.1.1 Wymiar: waluta

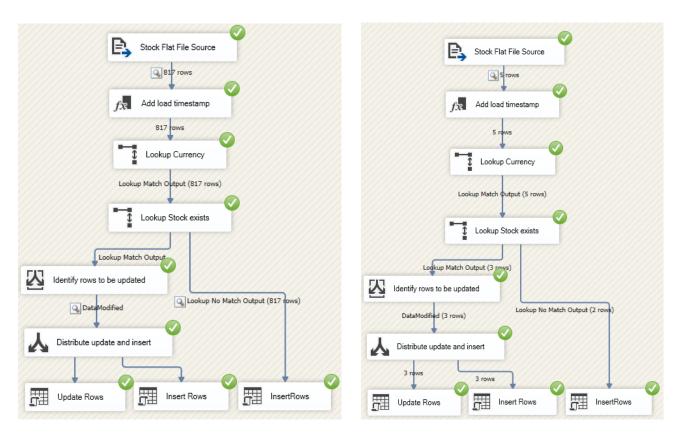




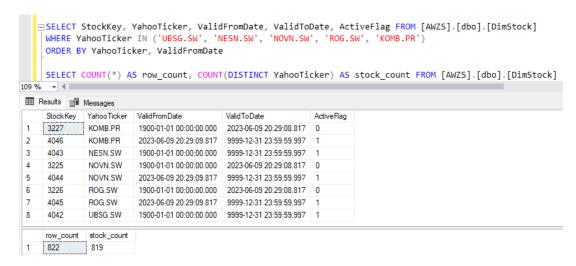
Rysunek 12: Ładowanie wymiaru waluty w SSIS (lewo) oraz zawartość tabeli w hurtowni (prawo)

Załadowanie sześciu obserwacji z pliku do wymiaru waluty było bezproblemowe.

11.1.2 Wymiar: spółka



Rysunek 13: Ładowanie wymiaru spółki w SSIS - plik oryginalny 817 wierszy oraz modyfikacja testująca SCD2 (5 wierszy, 2 nowe, 3 zaktualizowane)



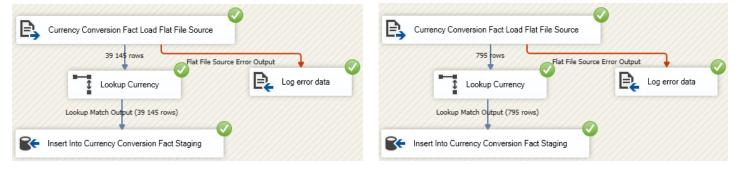
Rysunek 14: Weryfikacja zawartości tabeli w hurtowni po załadowaniu dwóch plików

Pierwsze ładowanie wymiaru spółki prawidłowo wstawiło 817 wierszy, natomiast drugie ładowanie zaktualizowało 3 wiersze, a wstawiło 2 nowe. Wynik zapytania SQL pokazuje, jak zadziałał mechanizm ładowania SCD2. Ostateczna liczba unikalnych spółek jest zgodnie z oczekiwaniami równa 819.

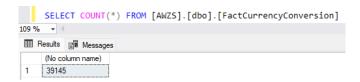
11.1.3 Fakt: kurs wymiany waluty

```
Iast_currency_conversion_fact_date.csv
                                                                     last_currency_conversion_fact_date.csv
      "CurrencyCode", "Date'
                                                                            "CurrencyCode", "Date"
       "CHF","2002-01-01"
                                                                            "CHF", "2022-12-31"
      "CZK","2002-01-01"
                                                                            "CZK", "2022-12-31"
 4
      "EUR", "2002-01-01"
                                                                            "EUR", "2022-12-31"
      "GBP", "2002-01-01"
                                                                            "GBP", "2022-12-31"
                                                                       5
      "USD", "2002-01-01"
                                                                            "USD", "2022-12-31"
```

Rysunek 15: Zawartość pliku last_currency_conversion_fact.csv po pierwszym (lewo) i drugim (prawo) uruchomieniu



Rysunek 16: Dwa uruchomienia procesu ładującego dane do tabeli faktów z kursami wymiany walut

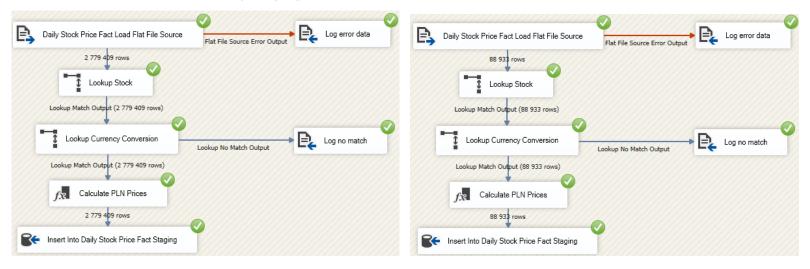


Rysunek 17: Ostateczna liczba wierszy w tabeli faktów z kursami wymiany walut

Miały miejsce dwa uruchomienia procesu ładującego dane do tabeli faktów z kursami wymiany walut. Pierwsze uruchomienie było inicjalizujące, tabela w hurtowni była pusta. Prawidłowo, do pliku last_currency_conversion_fact.csv

trafiła domyślna data 01.01.2002 r. i od tej daty zostały pobrane kursy wymiany walut aż do daty uruchomienia (10.06.2023 r.). Drugie uruchomienie miało symulować kolejną iterację danych. Z tego powodu z tabeli zostały wpierw usunięte rekordy od 01.01.2023 r. włącznie. Widzimy, że do pliku trafiły prawidłowe daty, a dane zostały prawidłowo doładowane od tego momentu. Finalnie, w tabeli znajduje się 39 145 wierszy.

11.1.4 Fakt: dzienne ceny akcji spółki



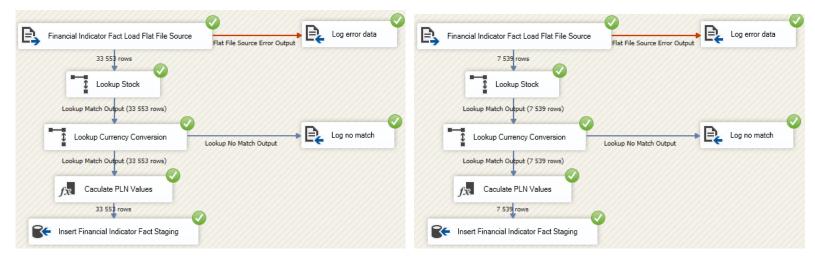
Rysunek 18: Dwa uruchomienia procesu ładującego dane do tabeli faktów z dziennymi cenami akcji spółek



Rysunek 19: Ostateczna liczba wierszy w tabeli faktów z dziennymi cenami akcji spółek

Ponownie, mają miejsce dwa uruchomienie: inicjalizujące oraz symulacja kolejnej iteracji danych (po usunięciu rekrodów od 01.01.2023 r.). W obu przypadkach wszystkie rekordy bezbłędnie zostały odczytane z pliku z pobranymi danymi oraz po transformacji trafiają do tabeli stagingowej. Ostateczna liczba wierszy w tabeli faktowej zgadza się z liczbą wierszy odczytanych z pliku.

11.1.5 Fakt: odczyt wskaźników finansowych spółki



Rysunek 20: Dwa uruchomienia procesu ładującego dane do tabeli faktów z odczytami wskaźników finansowych



Rysunek 21: Ostateczna liczba wierszy w tabeli faktów z odczytami wskaźników finansowych

Test funkcjonalny procesu ładującego dane do tabeli faktowej z odczytami wskaźników finansowych spółek przebiegł identycznie jak dla dwóch poprzednich tabel. Punktem "odcięcia" dla symulacji kolejnej iteracji danych była data 01.01.2020 r. Ten test również zakończył się pomyślnie.

11.2 Unikalność rekordów w wymiarze spółki

- Co jest testowane? unikalność klucza głównego (generowanego przez mechanizm *identity*) w wymiarze spółki; sprawdzenie, czy któraś spółka nie występuje kilka razy
- W jaki sposób jest testowane? odpowiednie polecenie SQL zliczające liczbę wierszy, liczbę unikalnych wartości klucza głównego, liczbę unikalnych spółek
- Jaki jest oczekiwany wynik? liczba wierszy jest równa liczbie unikalnych wartości klucza głównego; liczba unikalnych spółek jest równa 819 (tyle, ile było wstawionych)

```
COUNT(*) AS row_count,
COUNT(DISTINCT StockKey) AS no_unique_stock_key,
COUNT(DISTINCT StockName) AS stock_count
FROM [AWZS].[dbo].[DimStock]

109 % 

EM Results Messages

Tow_count no_unique_stock_key stock_count
1 822 822 819
```

Rysunek 22: Weryfikacja zawartości tabeli wymiaru spółki

11.3 Wiarygodność danych w tabeli faktów z dziennymi cenami akcji spółek

- Co jest testowane? wiarygodność danych w tabeli z dziennymi cenami akcji spółek, to jest czy ceny akcji i wolumen to liczby nieujemne oraz czy nie ma bardzo gwałtownych zmian (znacznego wzrostu lub spadku) między dwiema kolejnymi cenami zamknięcia
- W jaki sposób jest testowane? odpowiednie polecenia SQL: zliczające liczbę wierszy z ujemnymi wartościami wybranych kolumn oraz przedstawiające maksymalne zmiany procentowe pomiędzy dwiema kolejnymi cenami zamknięcia
- Jaki jest oczekiwany wynik? brak wierszy z ujemną wartością ceny akcji lub wolumenu; niewiele lub brak gwałtownych zmian pomiędzy dwoma kolejnymi cenami zamknięcia

```
□ SELECT

         COUNT(*) AS incorrect_records_count
    FROM [AWZS].[dbo].[FactStockDailyPrice]
        OpenPrice_LocalCurrency < 0
        HighPrice_LocalCurrency < 0
                                          OR
         LowPrice_LocalCurrency < 0
                                          OR
        ClosePrice LocalCurrency < 0
                                          OR
        OpenPrice_PLN < 0
                                          OR
        HighPrice PLN < 0
                                          OR
        LowPrice_PLN < 0
                                          OR
        ClosePrice_PLN < 0
                                          OR
        VolumeAmount < 0
109 %
Results Messages
    incorrect_records_count
```

(a) Sprawdzenie, czy ceny akcji spółek i wolumen to liczby nieujemne

```
⊟WITH tmp AS (
     SELECT
         StockKey.
         DateKey,
         ClosePrice LocalCurrency AS ClosePrice,
         LAG(ClosePrice_LocalCurrency, 1, NULL)
             OVER (PARTITION BY StockKey ORDER BY DateKey DESC) ClosePrice_NextDay
     FROM [AWZS].[dbo].[FactStockDailyPrice]
     WHERE ClosePrice_LocalCurrency > 0 -- błędy zaokrągleń
     SELECT
         DimStock.YahooTicker,
         DimDate.Date,
         ClosePrice.
         ClosePrice NextDay,
         ClosePrice_NextDay / ClosePrice - 1 AS Ratio
     FROM tmp
     JOIN DimStock ON DimStock.StockKey = tmp.StockKey
     JOIN DimDate ON DimDate.DateKey = tmp.DateKey
     WHERE ClosePrice_NextDay IS NOT NULL
     ORDER BY ClosePrice_NextDay / ClosePrice - 1 DESC
109 %
Results Messages
                Date
     Yahoo Ticker
                          ClosePrice ClosePrice NextDay
                                                   Ratio
     XELA
                2015-03-26
                         120.00
                                   117600.00
                                                   979.00
2
     XELA
                                   112800,00
                2015-03-13
                         120,00
                                                   939,00
 3
     LIN.DE
                2007-11-09 0.02
                                   17.50
                                                   874.00
     LIN.DE
                2007-08-13 0,05
                                   34,88
                                                   696,60
```

(b) Sprawdzenie, czy ceny akcji spółek ulegają gwałtownym zmianom

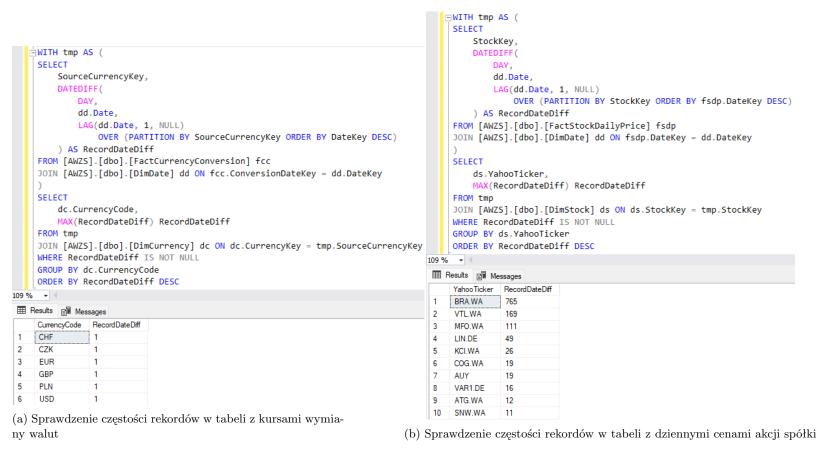
Rysunek 23: Sprawdzenie wiarygodności danych w tabeli faktów z dziennymi cenami akcji

Sprawdzenie, czy ceny akcji spółek i wolumen to liczby nieujemne, zakończyło się sukcesem - wynik polecenia na rysunku 23a jasno pokazuje, że nie ma niepoprawnych rekordów.

Fiaskiem natomiast okazał się test występowania nietypowych zmian cen akcji (rysynek 23b). Okazało się, że niespodziewanie bardzo dużo spółek doświadcza różnych dziwnych zmian cen akcji. Przykładem może byś Exela Technologies Inc. (XELA), której ceny akcji jednego dnia wzrosły prawie tysiąc razy. Po sprawdzeniu okazało się, że ten i inne przypadki to nie jest błąd procesu ETL lub hurtowni - takie dane były dostarczone przez Yahoo API. Z tego powodu, ten test nie może posłużyć do szukania błędów w procesie ETL lub hurtowni.

11.4 Wiarygodna częstotliwość danych w tabelach faktowych

- Co jest testowane? Wiarygodna częstotliwość danych w tabelach faktowych, to jest czy nie ma niespodziewanie długich przerw pomiędzy odczytami kursu waluty lub ceny akcji spółki
- W jaki sposób jest testowane? odpowiednie polecenia SQL, które przedstawią zestawienie najdłuższych
 przerw pomiędzy dwoma rekordami w tabelach faktowych dla jednej waluty czy spółki
- Jaki jest oczekiwany wynik? brak sytuacji, kiedy dwa odczyty cen akcji tej samej spółki czy kursu wymiany tej samej waluty są oddalone o więcej niż kilka dni;



W tabeli z kursami walut nie ma sytuacji, że dwa odczyty kursu wymiany tej samej waluty są oddalone o więcej niż jeden dzień. Natomiast w tabeli z dziennymi cenami akcji spółek znowu pojawia się problem. Okazuje się, że kilka spółek ma bardzo długie przerwy między kolejnymi odczytami cen. Znowu, to nie jest problem procesu ETL czy hurtowni - takie dane są dostarczane przez Yahoo API.

Literatura

- [1] nbp api. http://api.nbp.pl/.
- [2] ticker. https://pl.wikipedia.org/wiki/Symbol_gieldowy.
- [3] xbbg. https://pypi.org/project/xbbg/.
- [4] yfinance. https://pypi.org/project/yfinance.

Analiza techniczna - przegląd spółek

Spółki o największej zmianie ceny zamknięcia

Nazwa spółki	Cena zamknięcia	Zmiana % ▼
Blue Apron Holdings, Inc.	8,95	67,28%
Virgin Orbit Holdings, Inc.	0,05	25,00%
Nikola Corporation	0,70	14,75%
Auddia Inc.	0,43	13,15%
Signature Bank	0,13	8,33%
Kingsoft Cloud Holdings Limited	5,16	7,72%
Farfetch Limited	5,88	7,29%
Exela Technologies, Inc.	5,03	7,24%
MacroGenics, Inc.	5,56	6,92%
Cano Health, Inc.	1,41	6,01%
GameStop Corp.	22,68	5,78%
Varta AG	16,47	5,78%
Gaotu Techedu Inc.	3,13	5,74%
Aroundtown SA	1,15	5,50%
Ocado Group plc	386,60	5,48%
Ginkgo Bioworks Holdings, Inc.	1,85	5,11%
XPeng Inc.	8,87	4,84%
Bed Bath & Beyond Inc.	0,22	4,76%
Daqo New Energy Corp.	41,58	4,63%
Petróleo Brasileiro S.A Petrobras	13,76	4,63%
Gol Linhas Aéreas Inteligentes S.A.	4,17	4,25%

Spółki o największej zmianie wolumenu

Nazwa spółki	Wolumen	Zmiana % ▼
Auddia Inc.	2721200	8247,24%
Blue Apron Holdings, Inc.	26756600	2186,69%
Kingsoft Cloud Holdings Limited	8065200	902,76%
Shop Apotheke Europe N.V.	374460	712,03%
Peloton Interactive, Inc.	37815500	513,21%
BRF S.A.	4290000	475,68%
TAL Education Group	19229300	449,49%
Tencent Music Entertainment Group	38027300	418,83%
Aston Martin Lagonda Global Holdings plc	4229697	387,60%
ChargePoint Holdings, Inc.	33808800	292,01%
Companhia Brasileira De Distribuicao	652500	280,91%
Silvercorp Metals Inc.	2868800	278,17%
AMTD Digital Inc.	3009900	271,64%
Exela Technologies, Inc.	2393300	268,14%
Gaotu Techedu Inc.	4589200	257,19%
Dada Nexus Limited	7373800	246,51%
XPeng Inc.	24901800	237,39%
Blink Charging Co.	6489500	232,42%
Ambev S.A.	15043300	223,74%
17 Education & Technology Group Inc.	55500	215,34%
Ceconomy AG	1082311	208,12%
0.4.1.4.5.4.6	00500	105 000/

Analiza techniczna - analiza spółki



Yahoo Ticker

Branża

Analiza finansowa spółki

Nazwa spółki

11 bit studios S.A.

11 bit studios S.A.

Nazwa spółki

Poland

Kraj

Electronic Gaming & Multimedia

Branża

Communication Services

Sektor

11B PW Equity

Bloomberg Ticker

11B.WA

Yahoo Ticker

11 bit studios S.A. engages in production and sale of cross-platform video games worldwide. The company also operates publishing business, which publishes games produced by third-party development studios. 11 bit studios S.A. was incorporated in 2009 and is headquartered in Warsaw, Poland.

Podsumowanie biznesowe

Net Income

EBITDA

Revenue

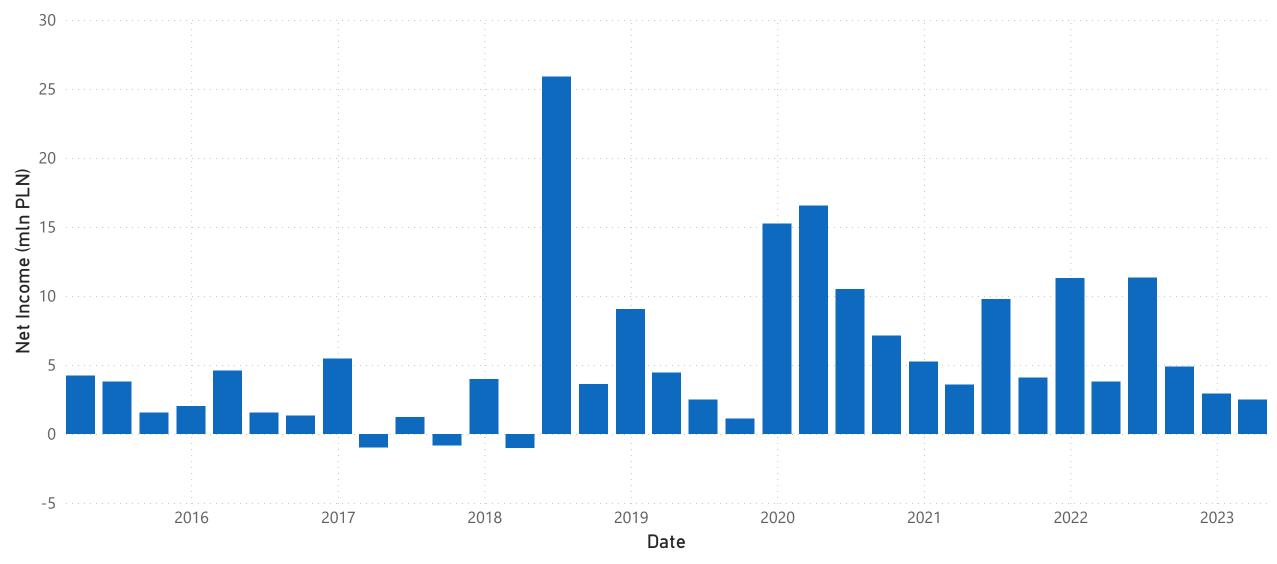
Current Ratio

Operating Margin

Profit Margin

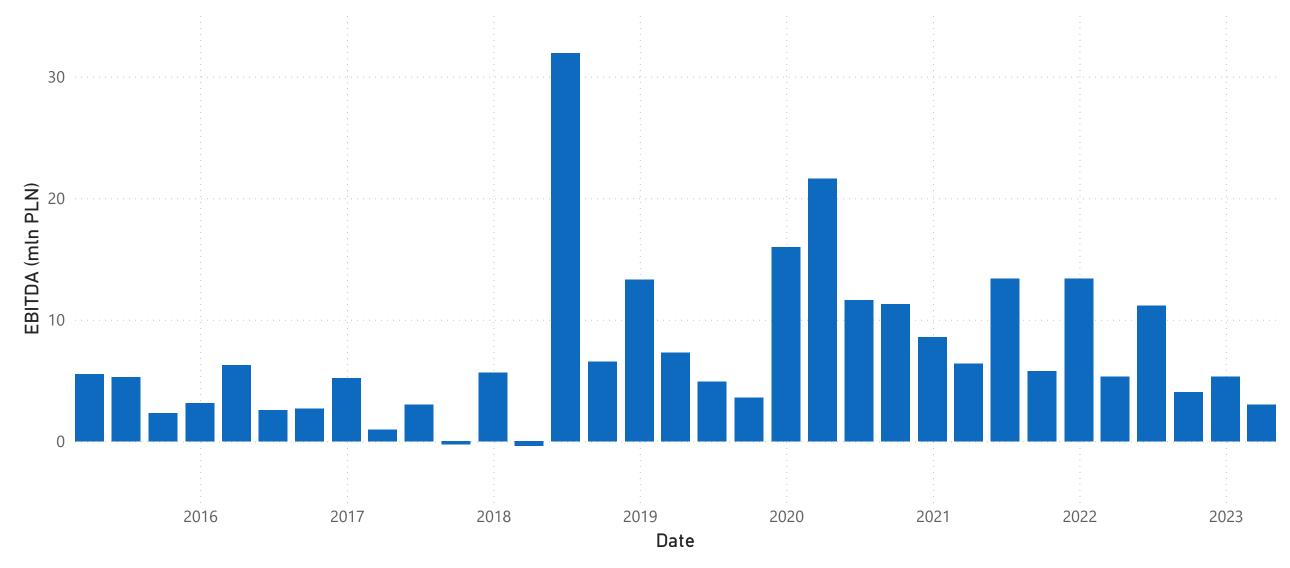
Return On Assets

Net Income



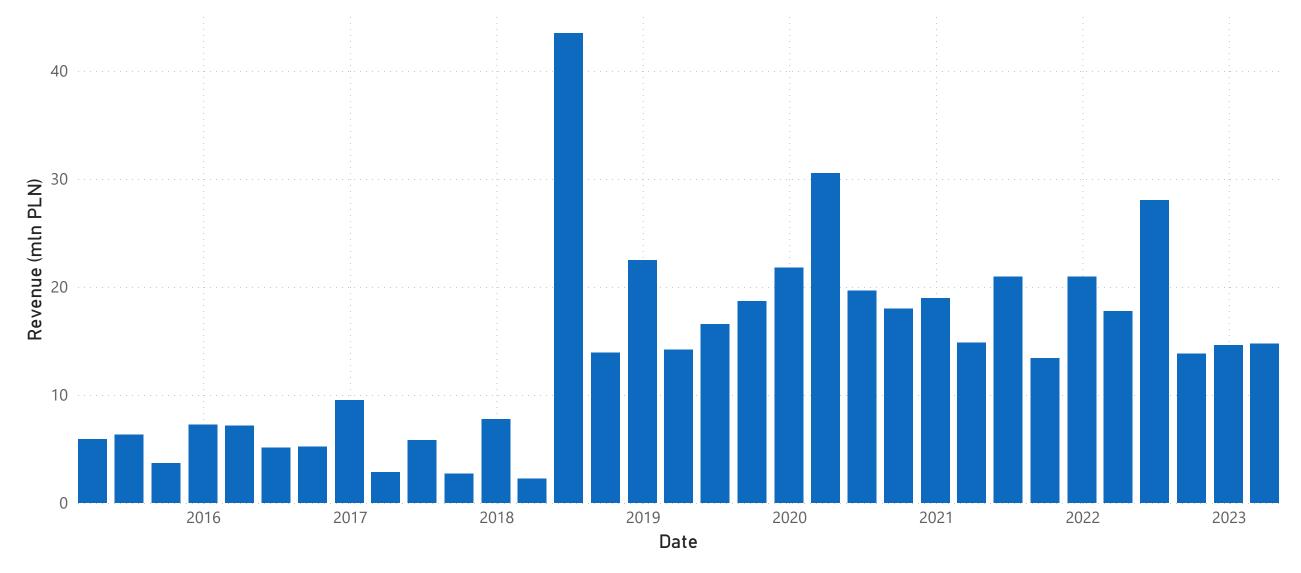
Net Income - a measure of a company's profitability that represents the amount of profit after deducting all expenses including taxes and interest from total revenue

EBITDA



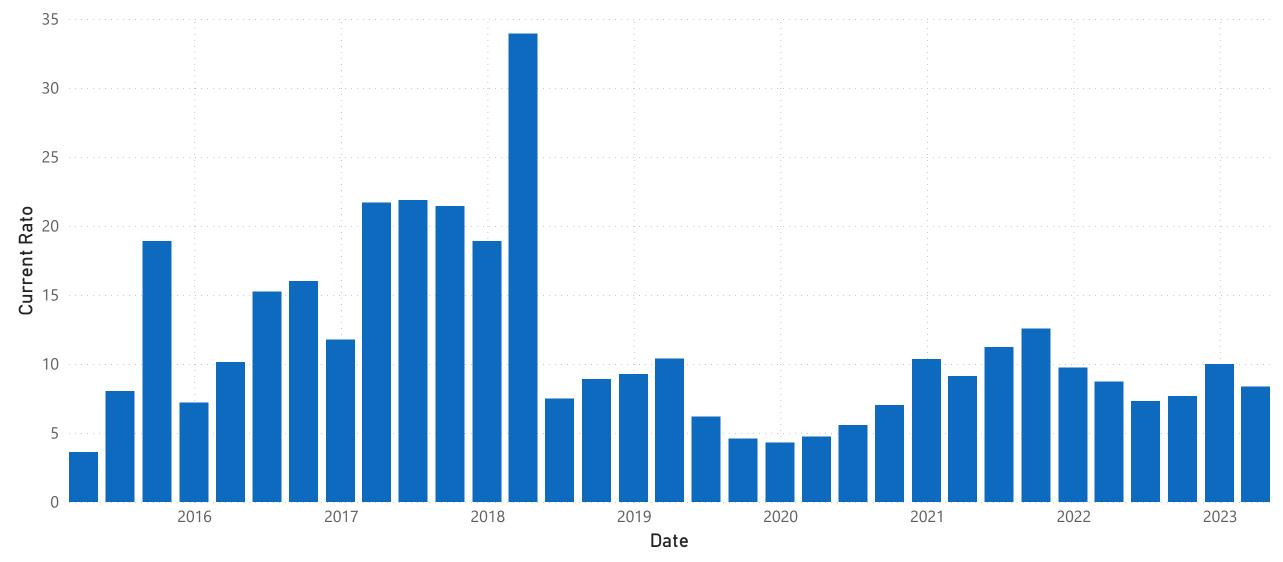
EBITDA - a measure of a company's financial performance that represents its earnings before interest taxes depreciation and amortization

Revenue



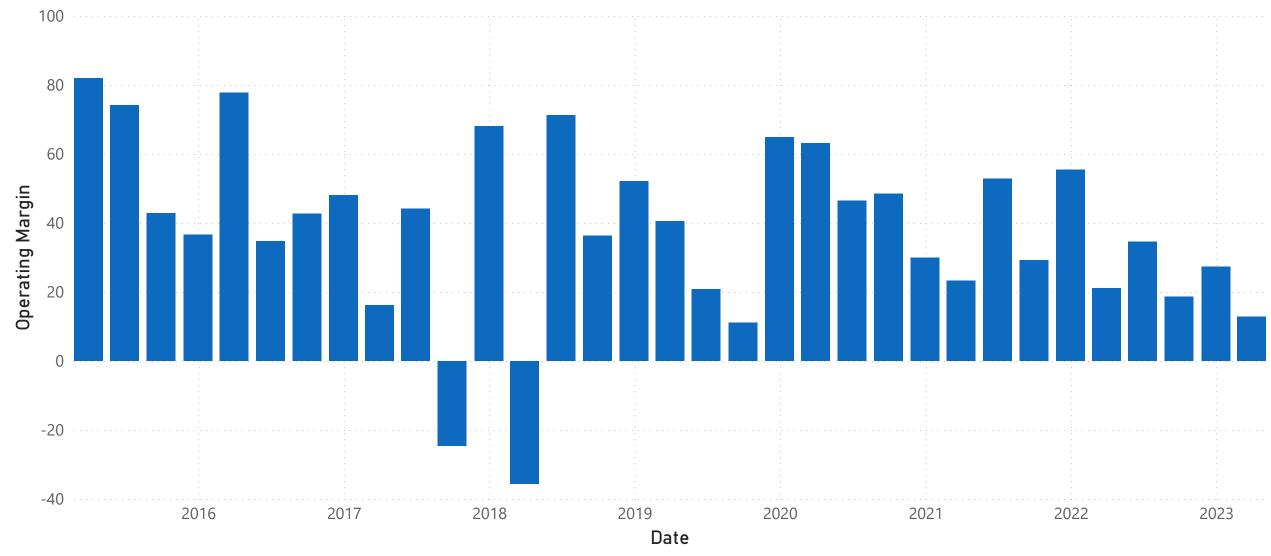
Revenue - total amount of money a company earns from the sale of its products or services during a given period

Current Ratio



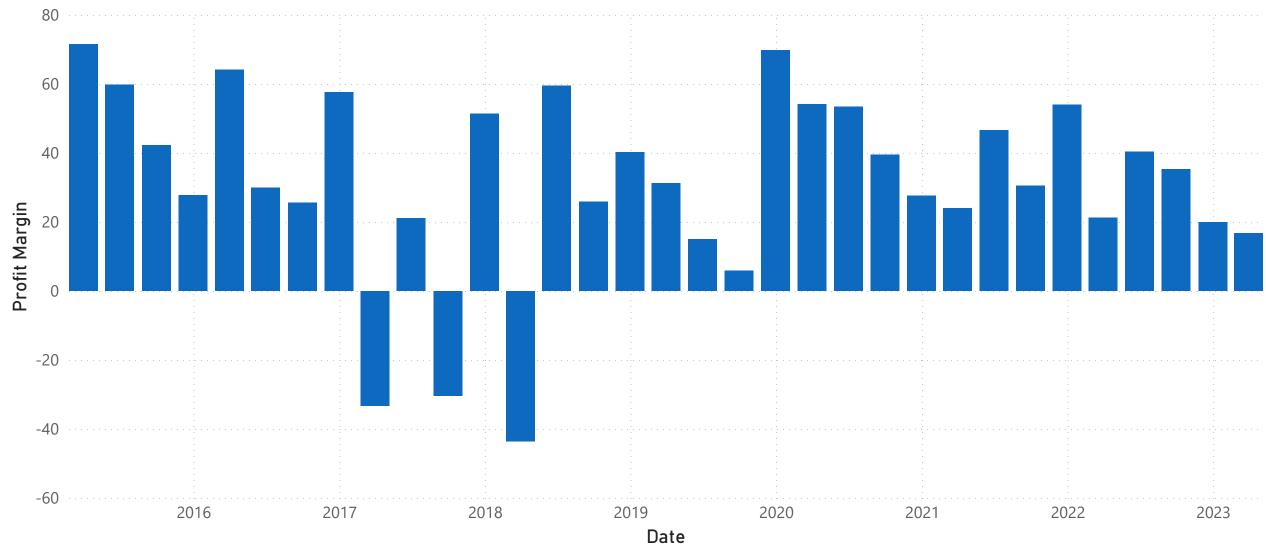
Current Ratio - a measure of a company's liquidity that compares its current assets to its current liabilities

Operating Margin



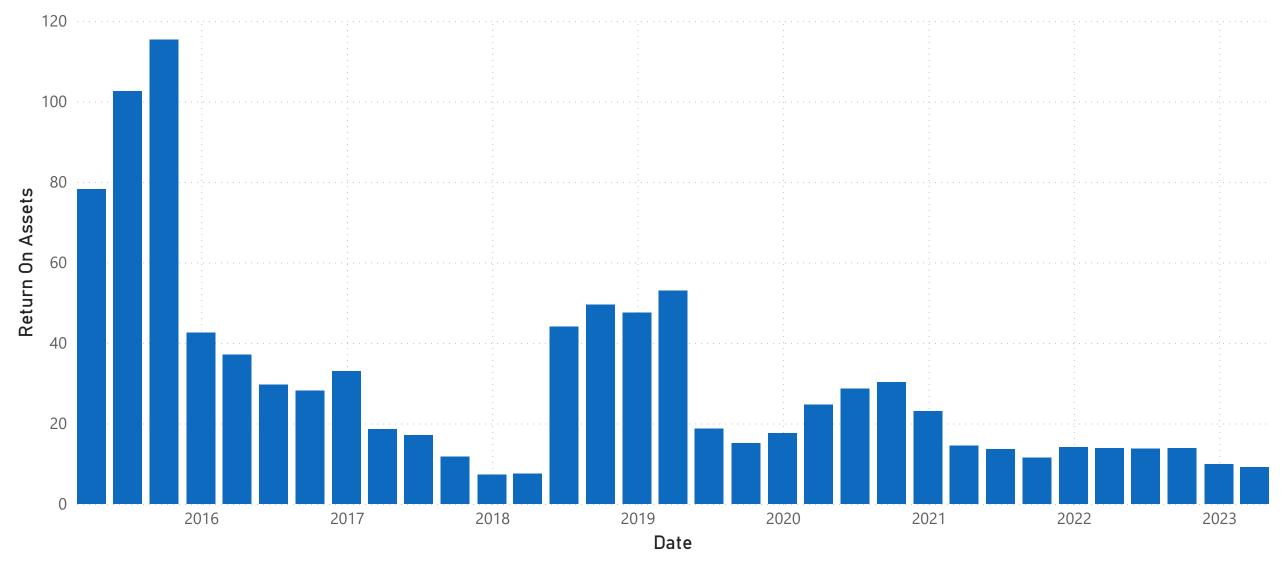
Operating Margin - a measure of a company's profitability that represents the percentage of revenue that remains after deducting all operating expenses

Profit Margin



Profit Margin - a measure of a company's profitability that represents the percentage of revenue that remains after deducting all expenses including taxes and interest

Return On Assets



Return On Assets - a measure of how efficiently a company uses its assets to generate profit