



Integracja Danych Airbnb i Giełdy Walut

Andrzejczuk Maja, Bielecki Piotr,
Hoang Laura



Cel projektu

Projekt zakłada stworzenie struktury do przechowywania, gromadzenia oraz integracji danych związanych z platformą **Airbnb** i **rynkiem walutowym**.

Potencjalne korzyści dla:

- **Uczestników Rynku Wynajmu Krótkoterminowego** - optymalizacja zarządzania ofertami oraz dostosowanie się do zmiennego otoczenia walutowego
- **Turystów** - ekonomiczne planowanie podróży, dostarczenie informacji pomocnych w wyborze atrakcyjnych lokalizacji biorąc pod uwagę zarówno koszty wynajmu jak i wynikające z kursów walutowych

Zbiory Danych

Zbiór danych Airbnb

- Dane pochodzą z API Airbnb.
- Kluczowe informacje: lokalizacja, daty pobytu, liczba gości, koszt.
- Odpowiedź w formie plików JSON z polem "results", które zawiera listę obiektów reprezentujących różne oferty noclegowe. Każdy z tych obiektów ma zestaw informacji.

Zbiór danych o walutach

- Dane pochodzą z API giełdy walut.
- Śledzenie bieżących i historycznych kursów walut.
- API zwraca stosunek wartości walut do jednej wybranej (np. USD, EUR).
- Zebrane metadane: timestamp i data pobrania informacji.

ADA	2.5984808614
AED	3.6724705473
AFN	73.5045710255
ALL	97.5421189163
AMD	402.0194040344
ANG	1.7865202566

Architektura systemu



Nifi - Automatyzacja przepływu danych

Hadoop - Przechowywanie danych dotyczących analizy walut i informacji z platformy Airbnb.

Hbase - Przechowywanie danych w czasie rzeczywistym w formacie tabelarycznym

PySpark - Analiza i przetwarzanie danych

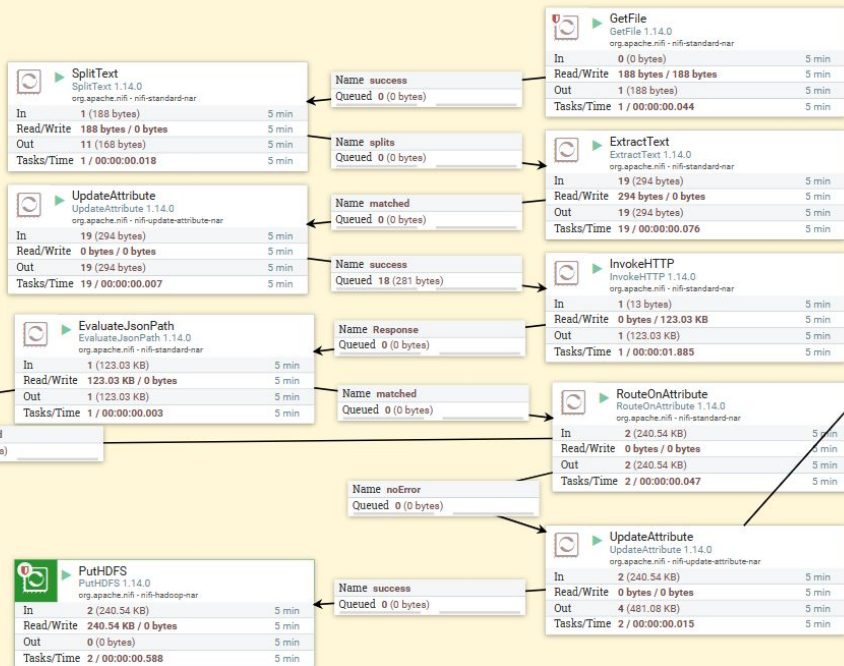


Proces przetwarzania danych

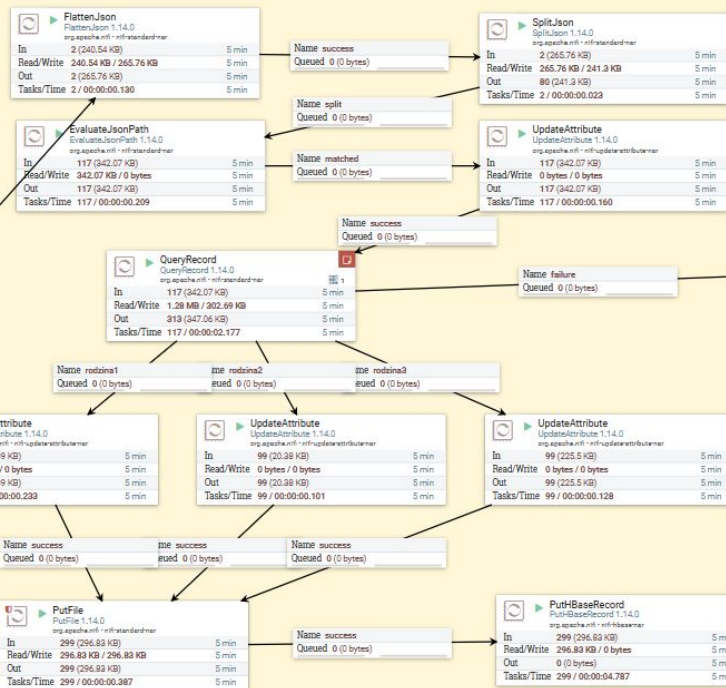


Proces przetwarzania danych z AirBNB API

Add files to HDFS



Add files to HBase



AirBNB API

Architektura systemu AirBNB API obejmuje pobieranie danych z API, ich transformację i zapis do HBase. Proces ten został zaprezentowany w NiFi, gdzie dane są podzielone na rodziny kolumn oraz uwzględniają daty pobrania informacji.

/home/vagrant/nifi/project/outputs/airbnb/				
Nazwa	Rozmiar	Zmodyfikowany	Prawa	Właściciel
		08.01.2024 12:34:11	rw-rw-r--x	vagrant
accomodation_003940.json	1 KB	07.01.2024 01:54:20	rw-r--r--	root
accomodation_004218.json	1 KB	07.01.2024 02:03:42	rw-r--r--	root
accomodation_005352.json	1 KB	07.01.2024 02:10:23	rw-r--r--	root
accomodation_005923.json	1 KB	07.01.2024 02:10:32	rw-r--r--	root
accomodation_010454.json	1 KB	07.01.2024 02:10:38	rw-r--r--	root
accomodation_010824.json	1 KB	07.01.2024 02:10:44	rw-r--r--	root
accomodation_012635.json	1 KB	07.01.2024 02:26:55	rw-r--r--	root
accomodation_150720.json	1 KB	07.01.2024 16:07:30	rw-r--r--	root
accomodation_151627.json	1 KB	07.01.2024 16:16:28	rw-r--r--	root
accomodation_151628.json	1 KB	07.01.2024 16:16:29	rw-r--r--	root
accomodation_151629.json	1 KB	07.01.2024 16:16:29	rw-r--r--	root
accomodation_151630.json	1 KB	07.01.2024 16:16:30	rw-r--r--	root
accomodation_153728.json	1 KB	07.01.2024 16:37:45	rw-r--r--	root
metadata_003955.json	2 KB	07.01.2024 02:10:50	rw-r--r--	root
metadata_004226.json	5 KB	07.01.2024 02:10:55	rw-r--r--	root
metadata_005356.json	3 KB	07.01.2024 02:10:59	rw-r--r--	root
metadata_005928.json	2 KB	07.01.2024 02:11:02	rw-r--r--	root
metadata_010458.json	4 KB	07.01.2024 02:11:07	rw-r--r--	root
metadata_010828.json	1 KB	07.01.2024 02:11:09	rw-r--r--	root
metadata_012641.json	2 KB	07.01.2024 02:27:08	rw-r--r--	root
metadata_150722.json	5 KB	07.01.2024 16:07:37	rw-r--r--	root
metadata_151627.json	2 KB	07.01.2024 16:16:27	rw-r--r--	root
metadata_151628.json	2 KB	07.01.2024 16:16:28	rw-r--r--	root
metadata_151629.json	2 KB	07.01.2024 16:16:30	rw-r--r--	root
metadata_151630.json	3 KB	07.01.2024 16:16:30	rw-r--r--	root
metadata_153733.json	3 KB	07.01.2024 16:37:44	rw-r--r--	root
money_003948.json	1 KB	07.01.2024 01:40:01	rw-r--r--	root
money_004223.json	1 KB	07.01.2024 01:42:41	rw-r--r--	root
money_005354.json	1 KB	07.01.2024 01:54:04	rw-r--r--	root
money_005925.json	1 KB	07.01.2024 01:59:38	rw-r--r--	root
money_010456.json	1 KB	07.01.2024 02:05:06	rw-r--r--	root

AirBNB API

Kolejnym krokiem jest analiza stworzonej tabeli 'airbnb' w HBase. Zauważamy, że tabela ta składa się z trzech rodzin kolumn: 'metadata', 'accomodation' oraz 'money'.

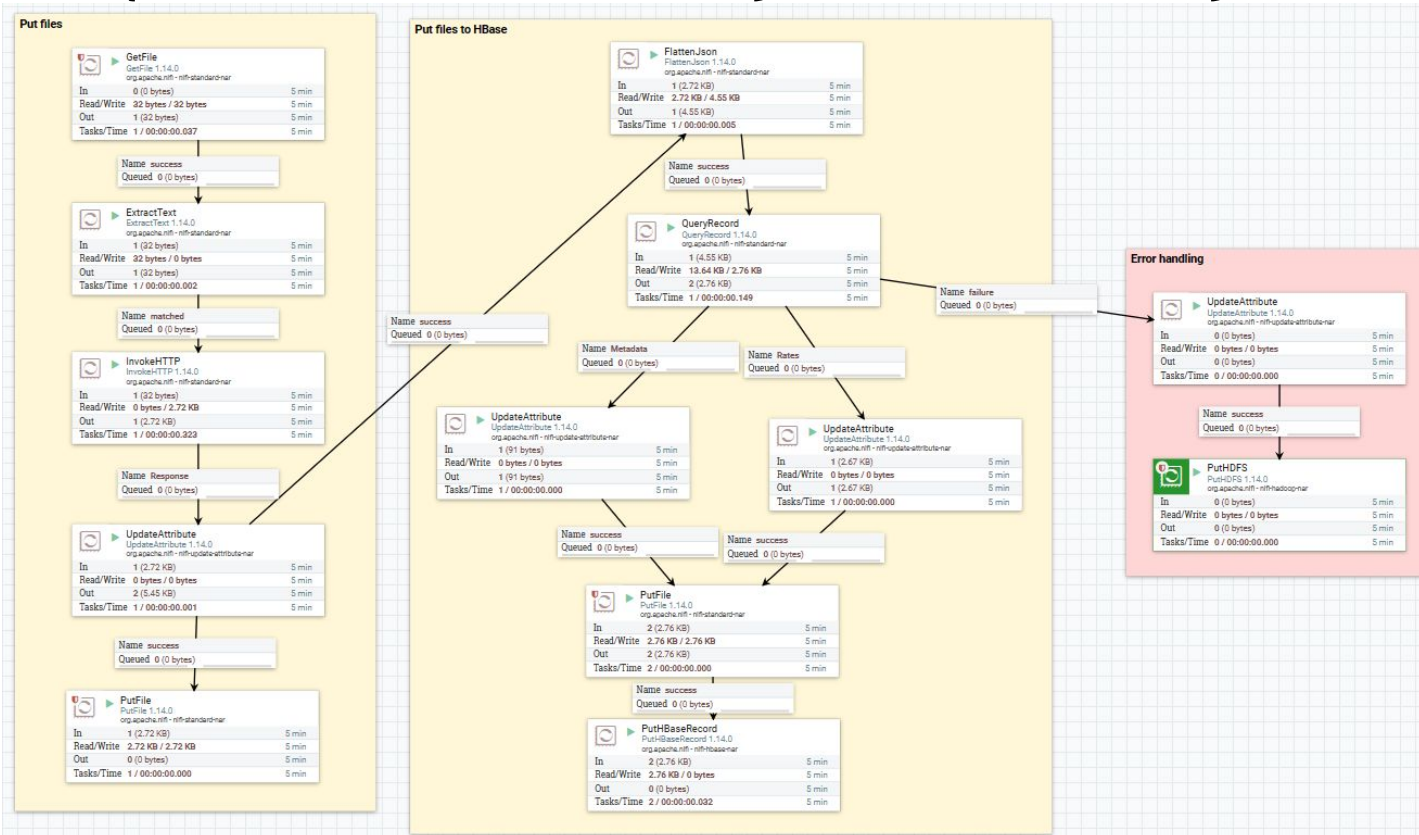
```
hbase(main):003:0> describe 'airbnb'
Table airbnb is ENABLED
airbnb
COLUMN FAMILIES DESCRIPTION
{NAME => 'accommodation', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}

{NAME => 'metadata', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}

{NAME => 'money', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}

3 row(s)
Quota is disabled
Took 0.0630 seconds
```


Proces przetwarzania danych z Currency API



Currency API

Architektura systemu Currency API obejmuje pobieranie danych z API, ich transformację i zapis do HBase. Proces ten został zaprezentowany w NiFi, gdzie dane są podzielone na rodziny kolumn oraz uwzględniają daty pobrania informacji.

/home/vagrant/nifi/project/outputs/currency2/				
Nazwa	Rozmiar	Zmodyfikowany	Prawa	Właściciel
↑ ..		08.01.2024 12:34:11	rw-rwxr-x	vagrant
metadata_20240108_131321.json	1 KB	08.01.2024 14:13:31	rw-r--r--	root
metadata_20240108_151200.json	1 KB	08.01.2024 16:12:00	rw-r--r--	root
metadata_20240108_151741.json	1 KB	08.01.2024 16:17:41	rw-r--r--	root
rates_20240108_131323.json	3 KB	08.01.2024 14:13:27	rw-r--r--	root
rates_20240108_151200.json	3 KB	08.01.2024 16:12:00	rw-r--r--	root
rates_20240108_151741.json	3 KB	08.01.2024 16:17:41	rw-r--r--	root

Currency API

Analizując stworzoną tabelę 'currency rates' w HBase, zauważamy dwie rodziny kolumn: 'Metadata' oraz 'Rates'.

```
hbase(main):002:0> describe 'currency_rates'
Table currency_rates is ENABLED
currency_rates
COLUMN FAMILIES DESCRIPTION
{NAME => 'Metadata', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}

{NAME => 'Rates', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}

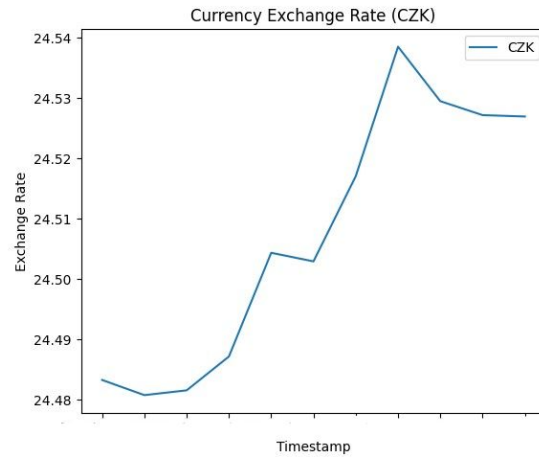
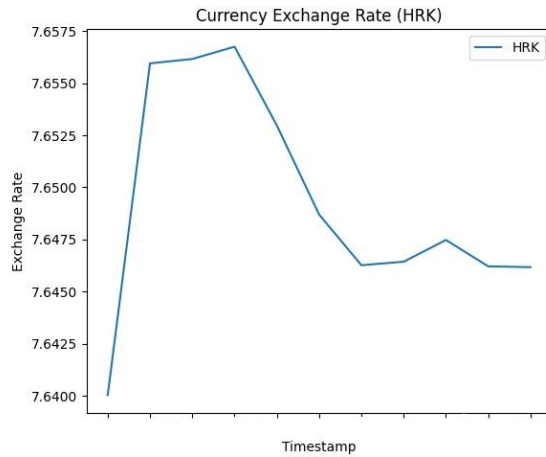
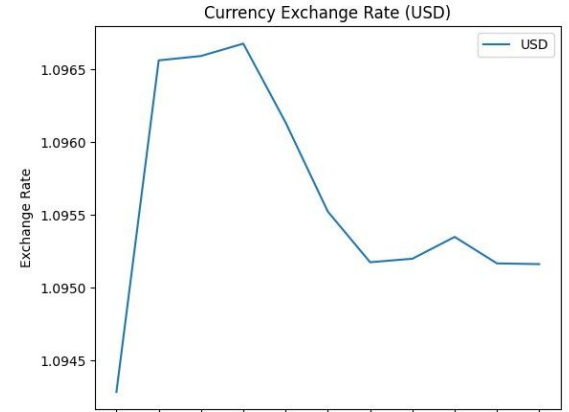
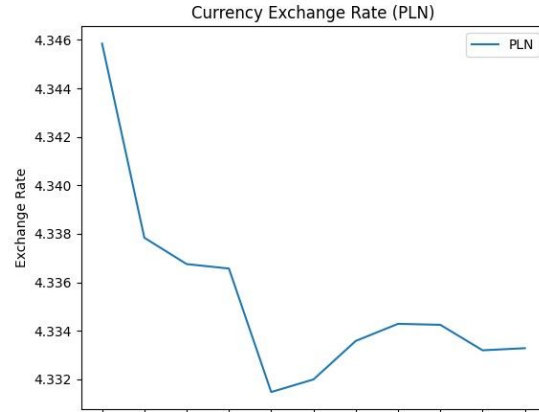
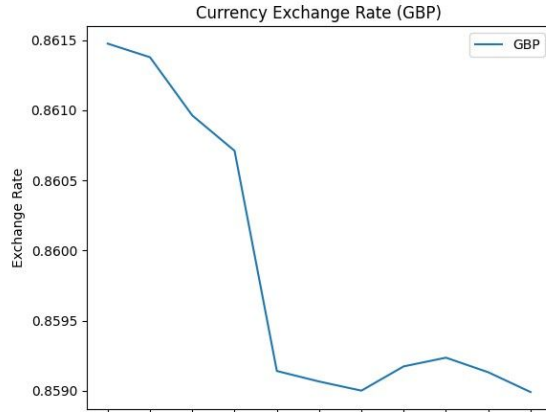
2 row(s)
Quota is disabled
Took 0.2132 seconds
```



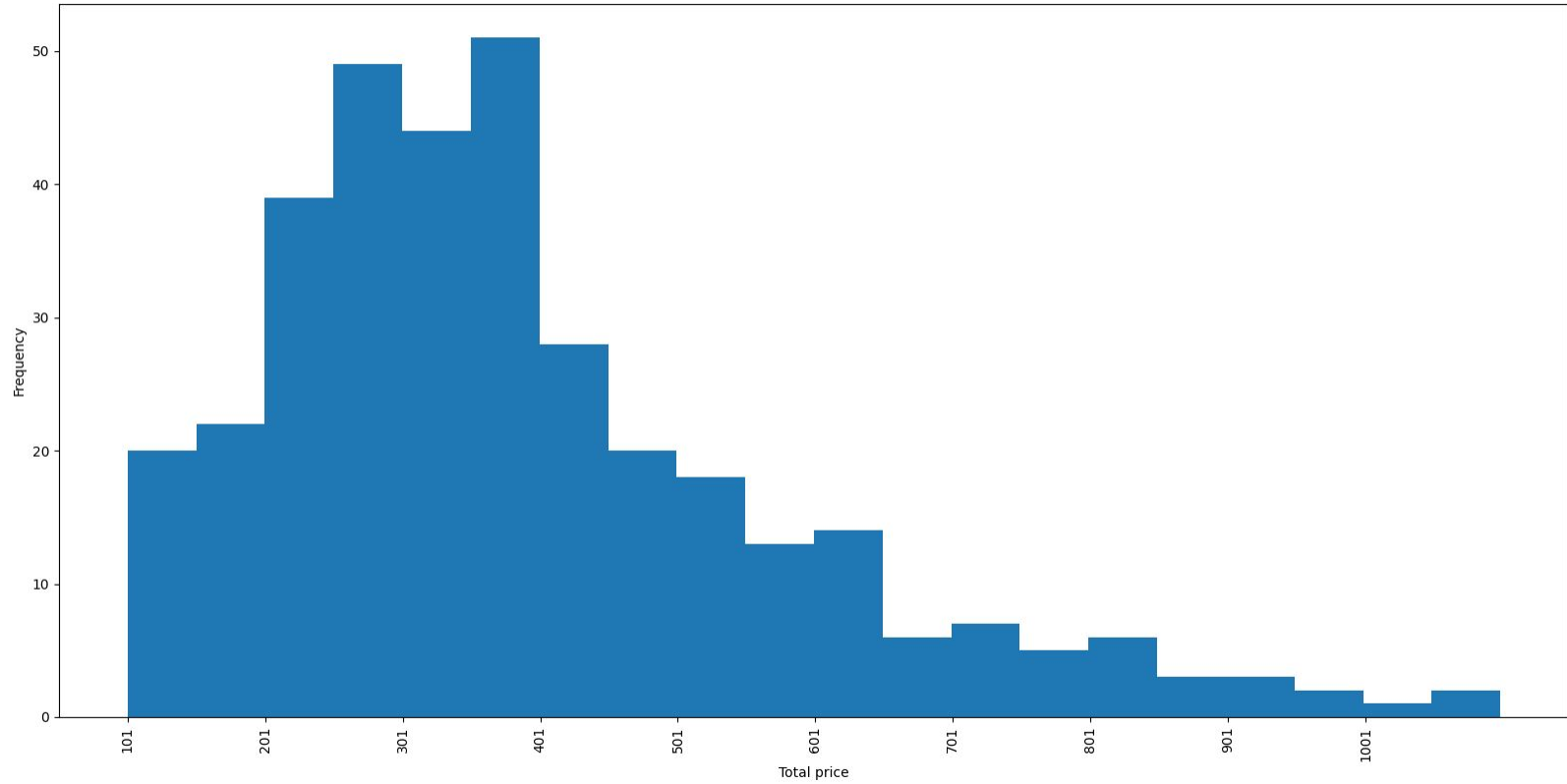
Warstwa prezentacyjna



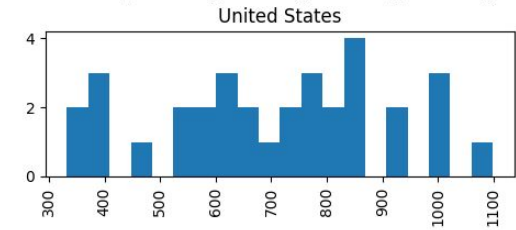
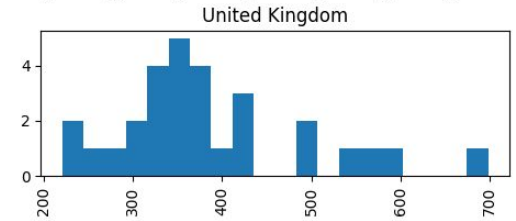
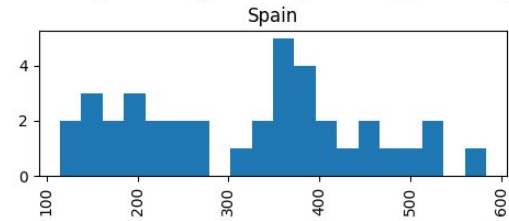
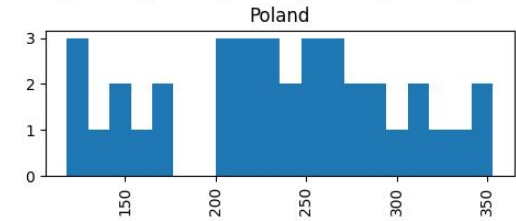
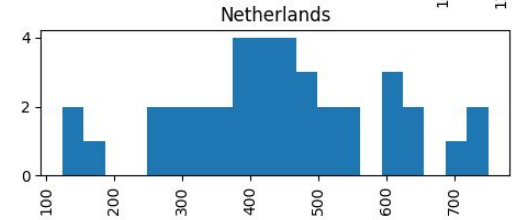
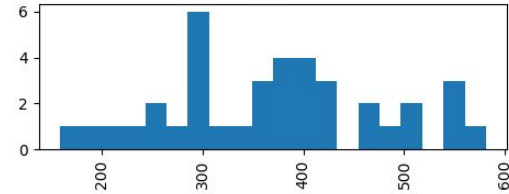
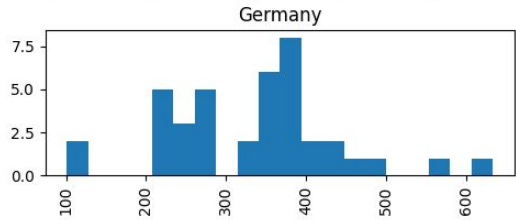
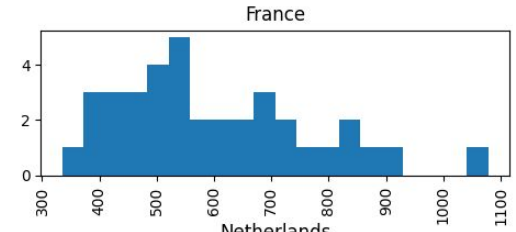
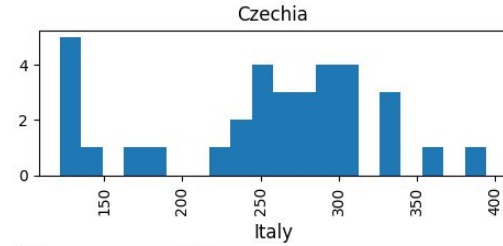
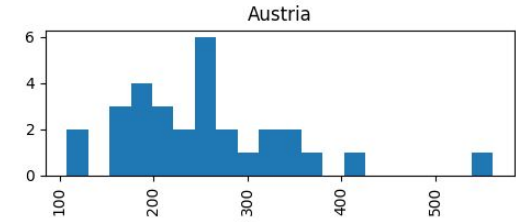
Currency stock changes



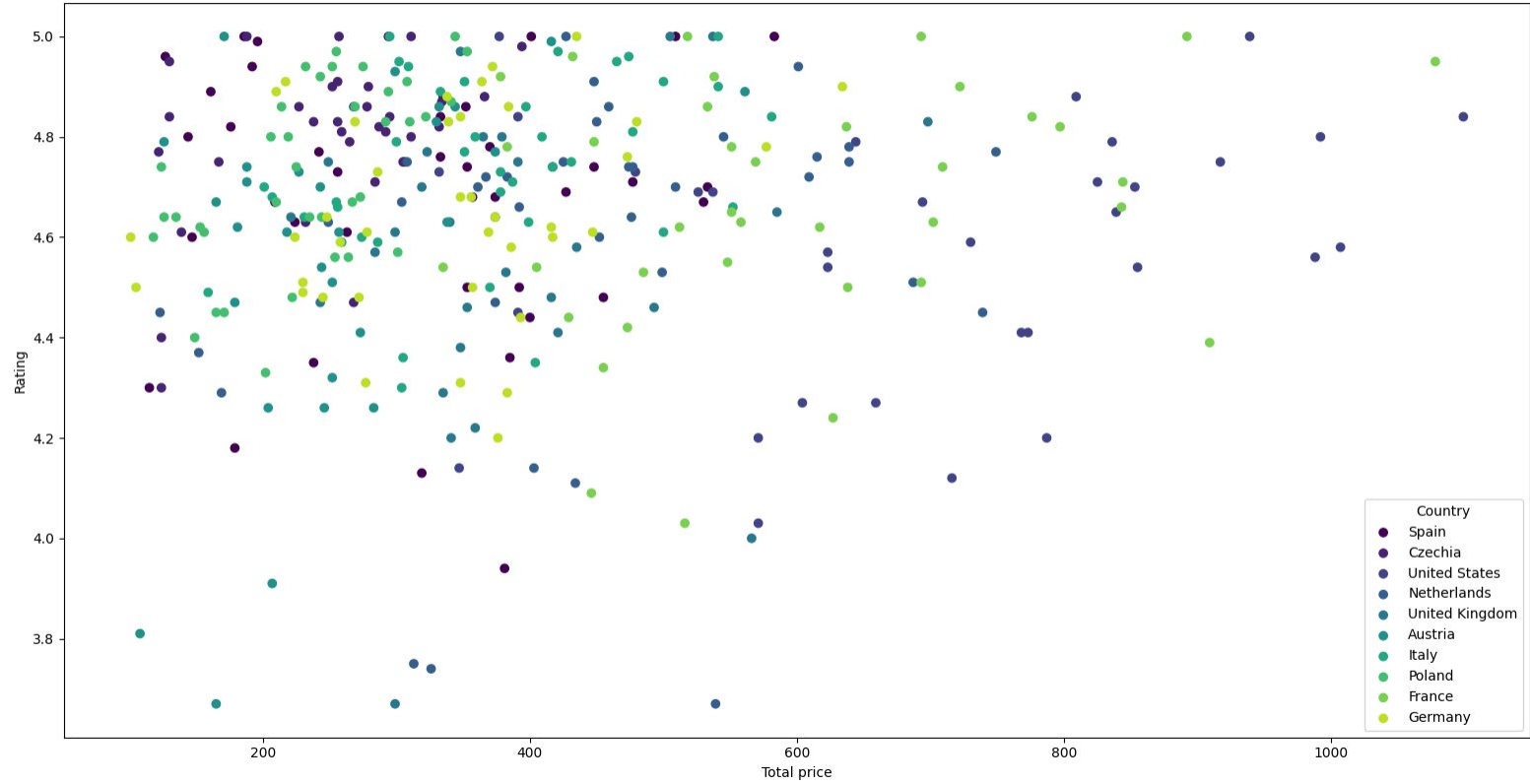
Histogram of total accommodation prices



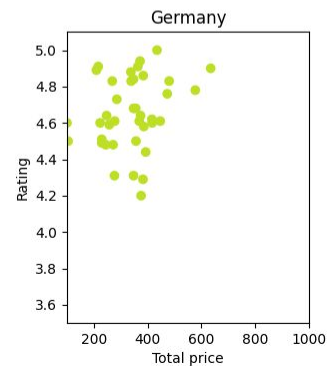
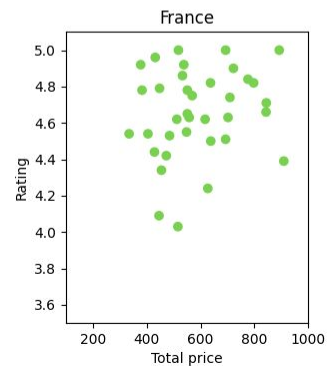
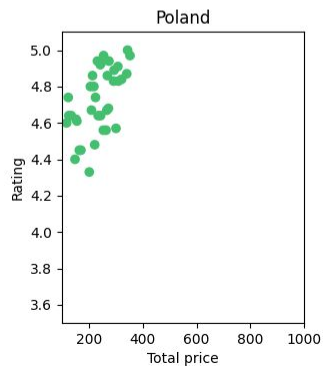
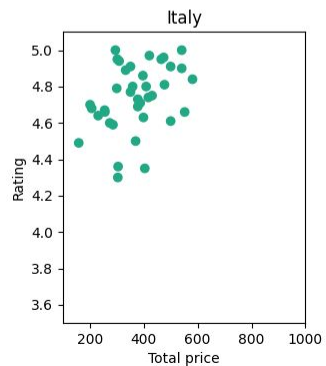
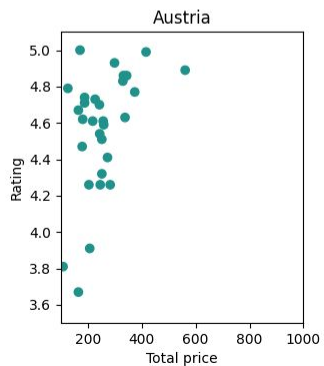
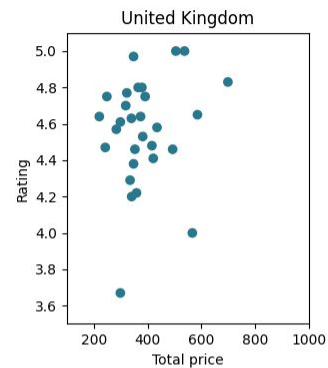
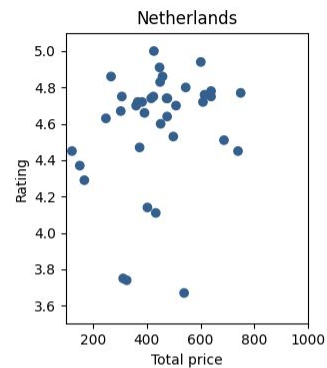
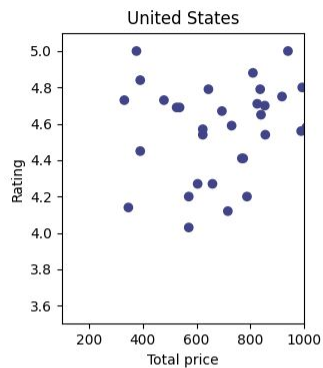
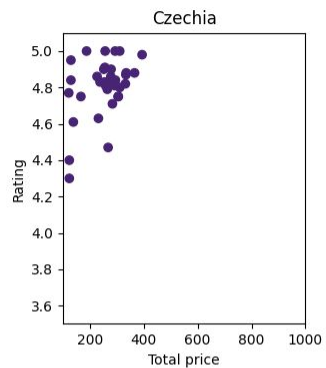
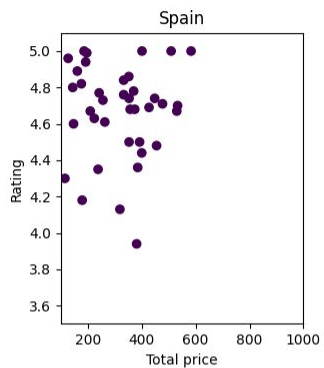
Histograms of total prices by countries



Comparison of Accommodation Prices to Ratings



Comparison of Accommodation Prices to Ratings Across Countries



Testy funkcjonalne

Przeprowadziliśmy szereg testów w celu zweryfikowania poprawności działania systemu. Skoncentrowaliśmy się na analizie wyników zapytań, sprawdzając ich dokładność i spójność.

- Badaliśmy struktury danych, upewniając się, że są zgodne z oczekiwaniami.
- Dokładnie weryfikowaliśmy outputy plików uzyskane na różnych etapach przetwarzania danych.
- Testy obejmowały porównanie danych przed i po różnych etapach, aby wykryć ewentualne błędy w przetwarzaniu.
- Cały proces miał na celu zapewnienie, że system działa zgodnie z oczekiwaniami, a dane są wiarygodne i zgodne z założeniami projektowymi.

Podsumowanie

Stworzony system do przechowywania, gromadzenia oraz integracji danych związanych z platformą AirBNB:

- Opiera się na rodzinie Apache.
- Przepływ danych został skonstruowany w sposób automatyczny, zapewniając dostępność danych do analizy.

Wpływ na podejmowanie decyzji biznesowych:

- System umożliwia monitorowanie ofert noclegowych oraz analizę dynamicznych zmian kursów walut.
- Kluczowe wsparcie dla decydentów, umożliwiające trafione decyzje i efektywne zarządzanie zasobami, dzięki dostarczeniu wiarygodnych danych i narzędziu analizy biznesowej.