

Mirosław Bober

Analiza opóźnień i odwołań lotów w USA w 2015 roku

Spis treści

Wstęp	2
Cele analizy	2
Hipotezy i pytania badawcze	3
Hipotezy badawcze	3
Pytania badawcze	3
Źródła danych	4
Przygotowanie danych	5
Analiza wstępna zbioru danych	8
Analiza	11
Liczba odwołań i opóźnień lotów	11
Top 10 lotnisk według liczby opóźnień i odwołań	12
Najczęściej zakłócone linie	13
Najczęściej występujące opóźnienia wg pory roku i dnia	14
Różnice pomiędzy regionami	16
Połączenia krajowe i międzynarodowe	17
Wnioski	18
Bibliografia	20

Wstęp

W niniejszym raporcie skupiono się na analizie danych dotyczących opóźnień i odwołań lotów na terenie Stanów Zjednoczonych w roku 2015. Ruch lotniczy w USA jest jednym z największych na świecie, a wszelkie zakłócenia w jego funkcjonowaniu mają wpływ nie tylko na pasażerów, lecz także na gospodarkę i bezpieczeństwo.

Analiza danych została podzielona na analizę wstępną z narzędziem Python i analizę właściwą przeprowadzoną z wykorzystaniem narzędzia Power BI. Celem pracy było nie tylko zrozumienie skali i charakterystyki opóźnień i odwołań lotów, ale także przedstawienie tych danych w formie przejrzystych i funkcjonalnych dashboardów. Skupiono się na analizie przestrzennej, czasowej oraz kontekstowej w celu odpowiedzenia na pytania badawcze i sprawdzenie postawionych hipotez.

Cele analizy

- Przedstawienie statystyk dotyczących liczby opóźnionych i odwołanych lotów w USA w 2015 roku
- Określenie regionów gdzie występowały największe problemy z punktualnością
- Porównanie efektywności poszczególnych linii lotniczych pod względem punktualności
- Przeprowadzenie analizy sezonowej i tygodniowej opóźnień
- Wizualna prezentacja danych z podziałem na lokalizacje, miesiące, linie lotnicze i inne istotne cechy

Hipotezy i pytania badawcze

Hipotezy badawcze

- Najwięcej opóźnień występuje w miesiącach letnich z powodu większego natężenia ruchu
- Największy odsetek opóźnień mają loty poranne (pierwsze rotacje)
- Duże lotniska międzynarodowe notują więcej opóźnień niż lotniska regionalne
- Linie budżetowe mają większy odsetek odwołań lotów
- Loty wieczorne są bardziej podatne na kumulacyjne opóźnienia

Pytania badawcze

Celem analizy jest odpowiedź na następujące pytania badawcze:

- Ile lotów zostało opóźnionych i odwołanych w 2015 roku w USA?
- Które lotniska miały największą liczbę opóźnień i odwołań?
- Jakie linie lotnicze charakteryzowały się największym odsetkiem zakłóconych lotów?
- W jakich porach roku oraz dnia najczęściej występowały opóźnienia?
- Czy istnieją różnice w opóźnieniach pomiędzy regionami?
- Jaki był przeciętny czas opóźnienia w zależności od typu lotu (krajowy, międzynarodowy)?

Źródła danych

Dane użyte w projekcie pochodzą z publicznie dostępnego zbioru danych pt. „2015 Flight Delays and Cancellations” opublikowanego na platformie Kaggle (<https://www.kaggle.com/datasets/usdot/flight-delays>). Zbiór zawiera szczegółowe informacje o ponad 5 milionach lotów komercyjnych w USA w 2015 roku. Dane zawierają między innymi:

- Identyfikatory linii lotniczych, lotnisk, tras
- Czas planowany i rzeczywisty startu oraz lądowania
- Informacje o opóźnieniach (czas, przyczyna)
- Status lotu (wykonany, opóźniony, odwołany, przekierowany)

Dane zostały udostępnione na licencji otwartej, co umożliwia ich przetwarzanie i publikację wyników w niniejszym raporcie.

# YEAR Year of the Flight Trip	# MONTH Month of the Flight Trip	# DAY Day of the Flight Trip	# DAY_OF_WEEK Day of week of the Flight Trip	▲ AIRLINE Airline Identifier	# FLIGHT_NUMBER Flight Identifier	▲ TAIL_NUMBER Aircraft Identifier
 2015 2015	 1 12	 1 31	 1 7	WN 22% DL 15% Other (3681343) 63%	 1 9855	4898 unique values
2015	1	1	4	AS	98	N407AS
2015	1	1	4	AA	2336	N3KUAA
2015	1	1	4	US	840	N171US
2015	1	1	4	AA	258	N3HYAA
2015	1	1	4	AS	135	N527AS
2015	1	1	4	DL	806	N3730B
2015	1	1	4	NK	612	N635NK
2015	1	1	4	US	2013	N584UW
2015	1	1	4	AA	1112	N3LAAA
2015	1	1	4	DL	1173	N826DN
2015	1	1	4	DL	2336	N958DN
2015	1	1	4	AA	1674	N853AA
2015	1	1	4	DL	1434	N547US
2015	1	1	4	DL	2324	N3751B

Rysunek 1: Próbka danych ze zbioru 'flights.csv'

Przygotowanie danych

Po pobraniu danych w formacie CSV zbiór został zaimportowany do programu Power BI. Ze względu na specyfikę danych (pochodzenie z USA) konieczne było dostosowanie formatu zapisu liczb (np. zmiana kropki dziesiętnej na przecinek w wartościach liczbowych) oraz użycie pierwszego wiersza jako nagłówek tabeli.

YEAR	MONTH	DAY	DAY OF WEEK	AIRLINE	FLIGHT NUMBER	TAIL NUMBER	ORIGIN	AIRPORT	DESTINATION	AIRPORT	SCHEDULED DEPARTURE	DEPARTURE TIME	DEPARTURE DELAY	TAXI OUT	WHEELS OFF	SCHEDULED TIME ELAPSED	T	U					
2	2015	11	4	AS	98	N407AS	SEA	SEA	0055	2354	-11	21	0015	205	194	168	1448	0404	0.0430	0408	-22.0	0	
3	2015	11	4	AA	2336	N3KUAA	LAX	PBI	0100	0002	-8	12	0114	280	279	263	2330	0737	0.0750	0741	-9.0	0	
4	2015	11	4	US	840	N171US	SFO	CLT	0020	0018	-2	16	0034	286	293	266	2296	0800	11	0806	0811	5.0	0
5	2015	11	4	AA	258	N3HYAA	LAX	MIA	0020	0015	-5	15	0030	285	281	258	2342	0748	8	0805	0756	-9.0	0
6	2015	11	4	AS	135	N527AS	SEA	ANC	0025	0024	-1	11	0035	235	215	199	1448	0254	5	0320	0259	-21.0	0
7	2015	11	4	DL	806	N3730B	SFO	MSP	0025	0020	-5	18	0038	217	230	206	1589	0604	6	0602	0610	8.0	0
8	2015	11	4	NK	612	N635NK	LAS	MSP	0025	0019	-6	11	0030	181	170	154	1299	0504	5	0526	0509	-17.0	0
9	2015	11	4	US	2013	N584UW	LAX	CLT	0030	0044	14	13	0057	273	249	228	2125	0745	8	0803	0753	-10.0	0
10	2015	11	4	AA	1112	N3LAAA	SFO	DFW	0030	0019	-11	11	0036	195	193	173	1464	0509	3	0545	0532	-13.0	0
11	2015	11	4	DL	1173	N826DN	LAS	ATL	0030	0033	3	12	0045	221	203	186	1747	0651	5	0711	0656	-15.0	0
12	2015	11	4	DL	2336	N958DN	DEN	ATL	0030	0024	-6	12	0036	173	149	133	1199	0449	4	0523	0453	-30.0	0
13	2015	11	4	AA	1674	N853AA	LAS	MIA	0035	0027	-8	21	0048	268	268	238	2174	0746	7	0803	0753	-10.0	0
14	2015	11	4	DL	1434	N547UW	LAX	MSP	0035	0035	0	18	0053	214	210	188	1535	0601	4	0609	0605	-4.0	0
15	2015	11	4	DL	2324	N3751B	SLC	ATL	0040	0034	-6	18	0052	215	199	176	1590	0548	5	0615	0553	-22.0	0
16	2015	11	4	DL	2440	N651DL	SEA	MSP	0040	0039	-1	28	0107	189	188	166	1399	0553	4	0549	0557	8.0	0
17	2015	11	4	AS	108	N309AS	ANC	SEA	0045	0041	-4	17	0058	204	194	173	1448	0451	4	0509	0455	-14.0	0
18	2015	11	4	DL	1560	N3743H	ANC	SEA	0045	0031	-14	25	0056	210	200	171	1448	0447	4	0515	0451	-24.0	0
19	2015	11	4	UA	1197	N78448	SFO	IAH	0048	0042	-6	11	0053	218	217	199	1635	0612	7	0628	0619	-7.0	0
20	2015	11	4	AS	122	N413AS	ANC	PDX	0050	0046	-4	11	0057	215	201	187	1542	0504	3	0525	0505	-18.0	0
21	2015	11	4	DL	1670	N808DN	PDX	MSP	0050	0045	-5	9	0054	193	186	171	1426	0545	6	0603	0551	-12.0	0
22	2015	11	4	NK	520	N525NK	SEA	MCI	0055	0120	25	11	0131	162	143	128	1138	0505	4	0537	0543	6.0	0
23	2015	11	4	AA	371	N3GXAA	SEA	MIA	0100	0052	-8	30	0122	338	347	311	2724	0933	6	0938	0939	1.0	0
24	2015	11	4	NK	214	N632NK	LAS	DFW	0103	0102	-1	13	0115	147	147	128	1055	0523	6	0530	0529	-1.0	0
25	2015	11	4	AA	115	N3CATA	LAX	MIA	0105	0103	-2	14	0117	286	276	255	2342	0832	7	0851	0839	-12.0	0
26	2015	11	4	DL	1450	N671DN	LAS	MSP	0105	0102	-3	11	0113	183	163	150	1299	0543	2	0608	0545	-23.0	0
27	2015	11	4	UA	1545	N76517	LAX	IAH	0115	0112	-3	11	0123	183	175	156	1379	0559	8	0618	0607	-11.0	0
28	2015	11	4	AS	130	N457AS	FAL	SEA	0115	0107	-8	25	0132	213	218	186	1533	0538	7	0548	0545	-3.0	0

Rysunek 2: Dane 'flights.csv' przed obróbką

	A	B	C
1	IATA_CODE	AIRLINE	
2	UA,United Air Lines Inc.		
3	AA,American Airlines Inc.		
4	US,US Airways Inc.		
5	F9,Frontier Airlines Inc.		
6	B6,JetBlue Airways		
7	OO,Skywest Airlines Inc.		
8	AS,Alaska Airlines Inc.		
9	NK,Spirit Air Lines		
10	WN,Southwest Airlines Co.		
11	DL,Delta Air Lines Inc.		
12	EV,Atlantic Southeast Airlines		
13	HA,Hawaiian Airlines Inc.		
14	MQ,American Eagle Airlines Inc.		
15	VX,Virgin America		

Rysunek 3: Dane 'airlines.csv' przed obróbką

	A	B	C	D	E	F	G	H	I
1	IATA_CODE	AIRPORT	CITY	STATE	COUNTRY	LATITUDE	LONGITUDE		
2	ABE	Lehigh Valley International Airport	Allentown	PA	USA	40.65236	-75.44040		
3	ABI	Abilene Regional Airport	Abilene	TX	USA	32.41132	-99.68190		
4	ABQ	Albuquerque International Sunport	Albuquerque	NM	USA	35.04022	-106.60919		
5	ABR	Aberdeen Regional Airport	Aberdeen	SD	USA	45.44906	-98.42183		
6	ABY	Southwest Georgia Regional Airport	Albany	GA	USA	31.53552	-84.19447		
7	ACK	Nantucket Memorial Airport	Nantucket	MA	USA	41.25305	-70.06018		
8	ACT	Waco Regional Airport	Waco	TX	USA	31.61129	-97.23052		
9	ACV	Arcata Airport	Arcata/Eureka	CA	USA	40.97812	-124.10862		
10	ACY	Atlantic City International Airport	Atlantic City	NJ	USA	39.45758	-74.57717		
11	ADK	Adak Airport	Adak	AK	USA	51.87796	-176.64603		
12	ADQ	Kodiak Airport	Kodiak	AK	USA	57.74997	-152.49386		
13	AEX	Alexandria International Airport	Alexandria	LA	USA	31.32737	-92.54856		
14	AGS	Augusta Regional Airport	Augusta	GA	USA	33.36996	-81.96450		
15	AKN	King Salmon Airport	King Salmon	AK	USA	58.67680	-156.64922		
16	ALB	Albany International Airport	Albany	NY	USA	42.74812	-73.80298		
17	ALO	Waterloo Regional Airport	Waterloo	IA	USA	42.55708	-92.40034		
18	AMA	Rick Husband Amarillo International Airport	Amarillo	TX	USA	35.21937	-101.70593		
19	ANC	Ted Stevens Anchorage International Airport	Anchorage	AK	USA	61.17432	-149.99619		
20	APN	Alpena County Regional Airport	Alpena	MI	USA	45.07807	-83.56029		
21	ASE	Aspen-Pitkin County Airport	Aspen	CO	USA	39.22316	-106.86885		
22	ATL	Hartsfield-Jackson Atlanta International Airport	Atlanta	GA	USA	33.64044	-84.42694		
23	ATW	Appleton International Airport	Appleton	WI	USA	44.25741	-88.51948		
24	AUS	Austin-Bergstrom International Airport	Austin	TX	USA	30.19453	-97.66987		
25	AVL	Asheville Regional Airport	Asheville	NC	USA	35.43619	-82.54181		
26	AVP	Wilkes-Barre/Scranton International Airport	Wilkes-Barre/Scranton	PA	USA	41.33815	-75.72427		
27	AZO	Kalamazoo/Battle Creek International Airport	Kalamazoo	MI	USA	42.23488	-85.55206		
28	BDL	Bradley International Airport	Windsor Locks	CT	USA	41.93887	-72.68323		

Rysunek 4: Dane 'airports.csv' przed obróbką

Plik flights.csv:

- Kolumny DIVERTED oraz CANCELLED zostały przekonwertowane z wartości liczbowych 0/1 na logiczne false/true, co ułatwiło interpretację binarnych zdarzeń w raportach.
- W przypadku brakujących wartości w kolumnie CANCELLATION_REASON wprowadzono domyślną wartość No cancellation.
- Puste (null) wartości w kolumnach dotyczących opóźnień (AIR_SYSTEM_DELAY, SECURITY_DELAY, AIRLINE_DELAY, LATE_AIRCRAFT_DELAY, WEATHER_DELAY) zostały zastąpione zerami, aby zapewnić spójność typów danych i poprawność obliczeń.

Plik airports.csv:

- Kolumny LATITUDE i LONGITUDE zostały przekonwertowane na format liczby dziesiętnej poprzez zmianę ustawień regionalnych z USA na polskie. Dzięki temu wartości geograficzne zostały poprawnie zinterpretowane w Power BI jako liczby zmiennoprzecinkowe.

Plik airlines.csv:

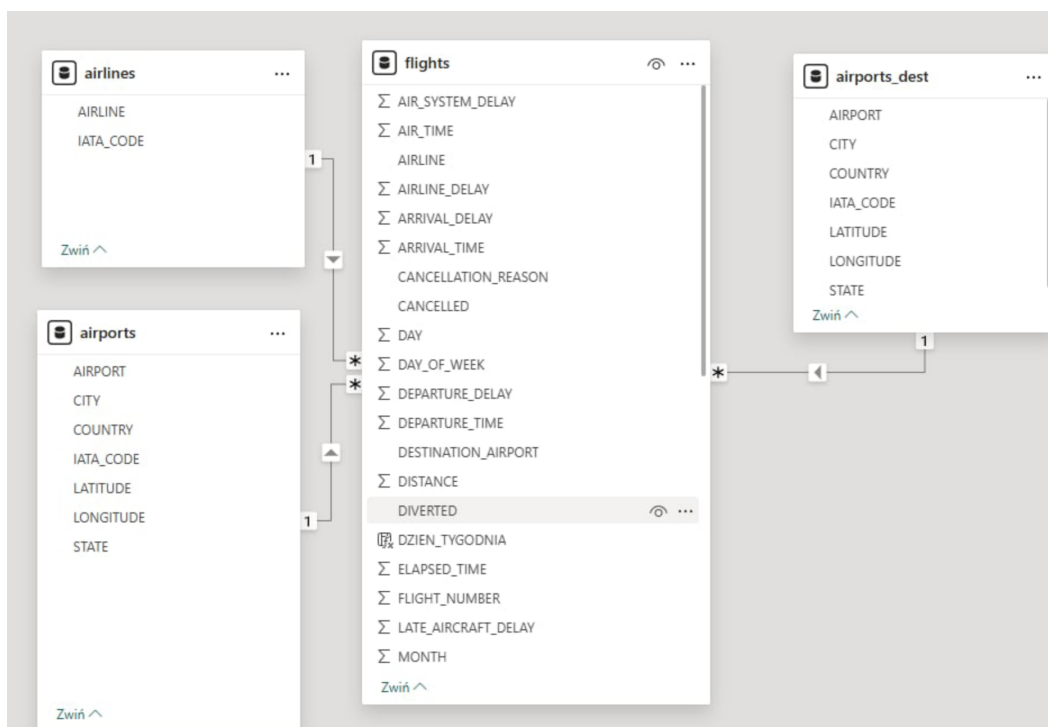
- Plik nie wymagał transformacji – wszystkie dane były poprawne i gotowe do użycia bez dodatkowej obróbki.

Relacje pomiędzy tabelami:

W celu zapewnienia poprawnej analizy danych w Power BI utworzono model relacyjny łączący tabelę **flights** z tabelami pomocniczymi **airlines** oraz **airports**. Kluczowe relacje zrealizowano na podstawie następujących pól:

- **flights(AIRLINE)** → **airlines(IATA_CODE)**,
- **flights(ORIGIN_AIRPORT)** → **airports(IATA_CODE)**,
- **flights(DESTINATION_AIRPORT)** → **airports_dest(AIRPORT)**.

Z uwagi na ograniczenia Power BI, które pozwala na istnienie tylko jednej aktywnej relacji między daną parą tabel, konieczne było zduplikowanie tabeli **airports** i nadanie jej nowej nazwy: **airports_dest**. Dzięki temu możliwe było osobne połączenie kolumny **ORIGIN_AIRPORT** z oryginalną tabelą lotnisk oraz kolumny **DESTINATION_AIRPORT** z jej kopią.



Rysunek 5: Schemat relacji tabel



Rysunek 6: Relacje połączeń

Analiza wstępna zbioru danych

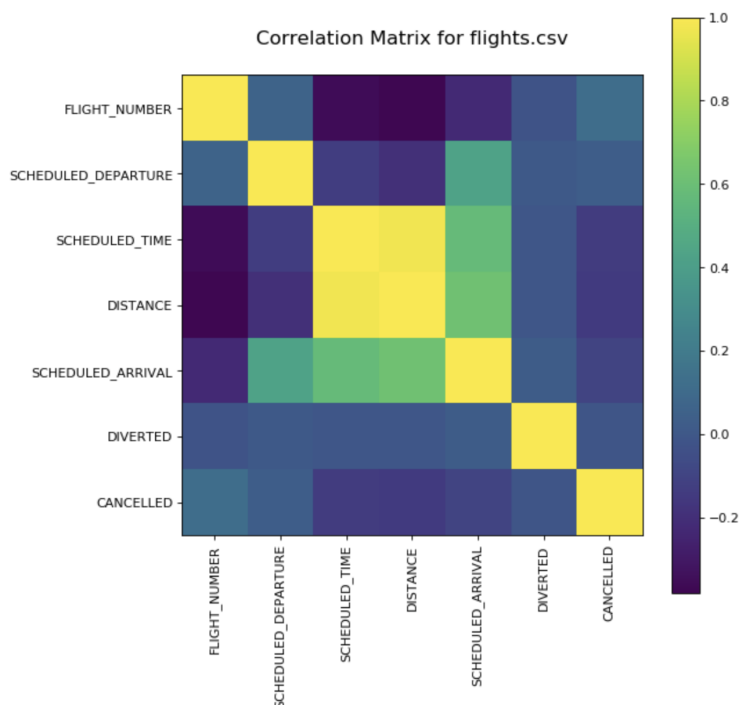
Po zakończeniu etapu obróbki i transformacji danych, a przed przystąpieniem do właściwej wizualizacji w Power BI, przeprowadzono analizę wstępną mającą na celu ogólne rozpoznanie zależności występujących w zbiorze. Celem tego kroku było wczesne zidentyfikowanie ukrytych powiązań między zmiennymi, wykrycie ewentualnych odstających obserwacji oraz zrozumienie struktury danych liczbowych.

Analiza wstępna została wykonana z wykorzystaniem środowiska Python i pakietów analizy danych takich jak `pandas` oraz `seaborn`.

Na poniższym wykresie zaprezentowano macierz korelacji zmiennych liczbowych występujących w zbiorze `flights.csv`. Najsilniejsze dodatnie korelacje występują pomiędzy:

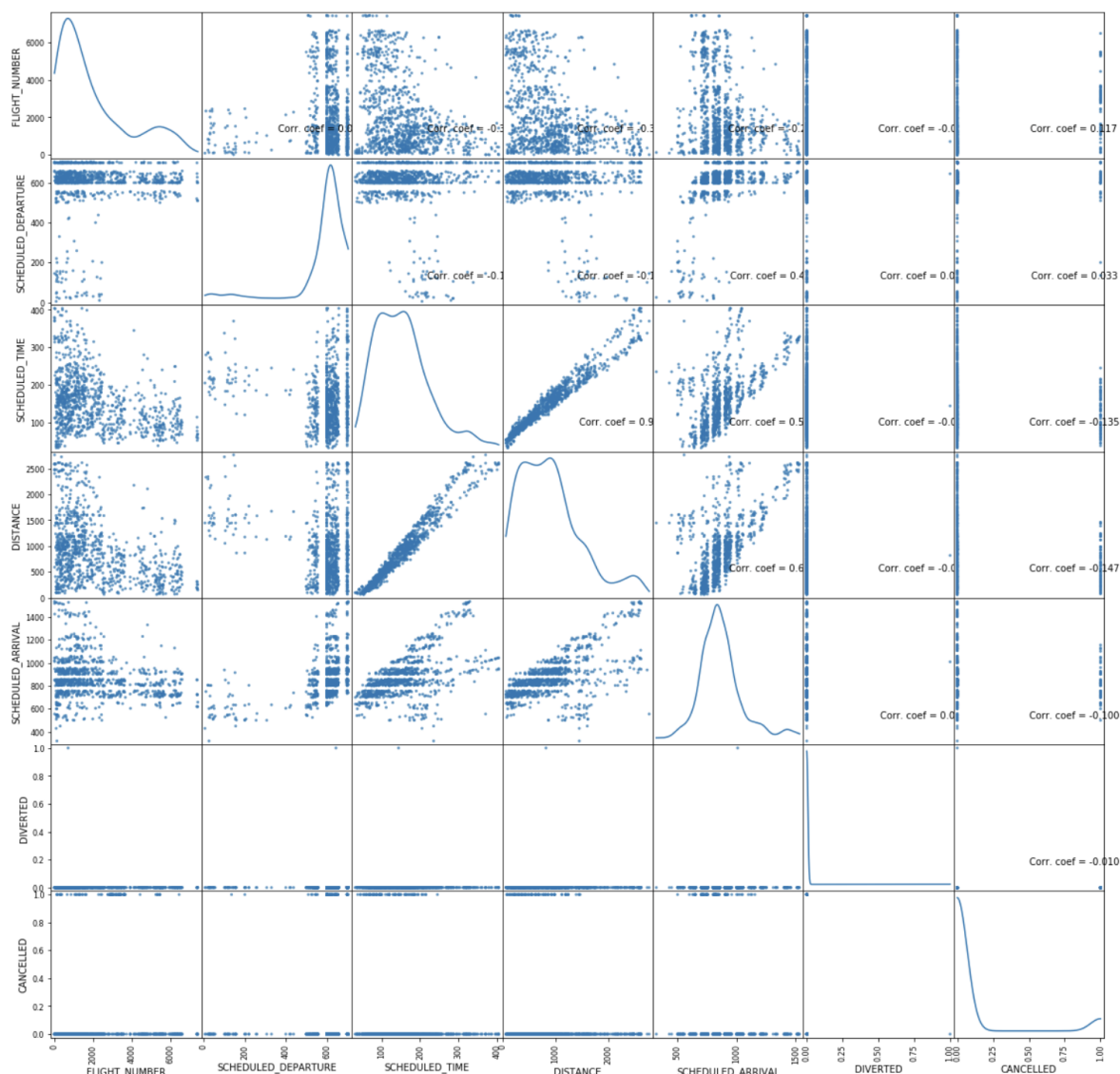
- `SCHEDULED_TIME`, `DISTANCE` i `SCHEDULED_ARRIVAL` – co odzwierciedla logiczne powiązanie między długością lotu, jego dystansem oraz czasem trwania.

Zmienna `CANCELLED` nie koreluje znacząco z żadnym z atrybutów numerycznych, co może wskazywać na losowy lub zewnętrzny charakter decyzji o odwołaniu lotu (np. warunki pogodowe, sytuacje operacyjne). Analogicznie, `DIVERTED` również nie wykazuje silnych zależności z pozostałymi kolumnami.



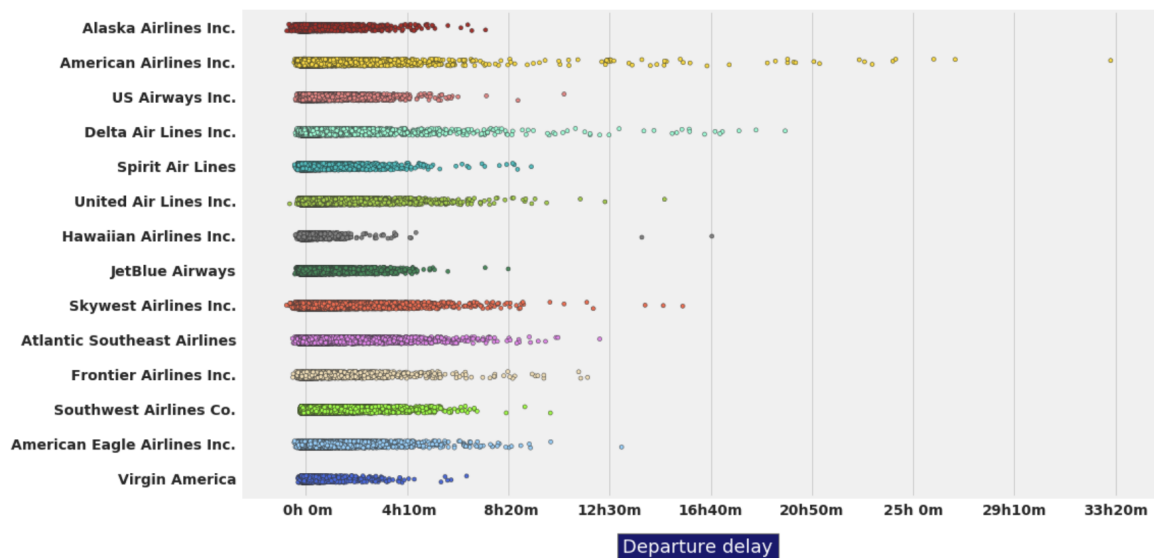
Rysunek 7: Macierz korelacji

Dodatkowo, w celu głębszego wglądu w rozkład danych i ewentualne zależności między zmiennymi liczbowymi, wygenerowano macierz wykresów rozrzutu (ang. *pair-plot*) z naniesionymi gęstościami oraz współczynnikami korelacji. Wykres ten pozwala zaobserwować nie tylko siłę korelacji, ale również rozrzut i kształt rozkładów każdej z cech względem pozostałych.



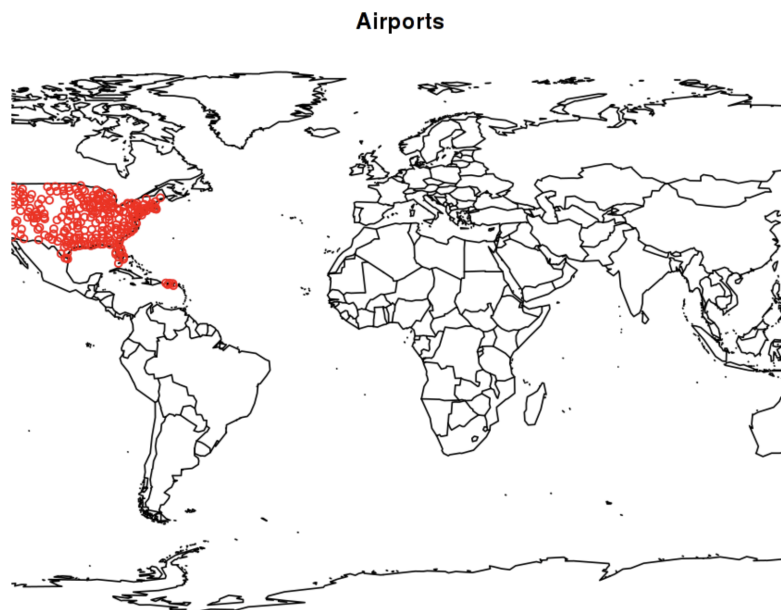
Rysunek 8: Wykres rozrzutu cech

Na wykresie zauważalne są silne liniowe zależności pomiędzy zmiennymi **SCHEDULED_TIME**, **DISTANCE** oraz **SCHEDULED_ARRIVAL**, co jest zgodne z wcześniejszą analizą macierzy korelacji. Pozostałe pary cech wykazują bardzo słabe lub zerowe współczynniki korelacji, co sugeruje brak istotnych zależności liniowych pomiędzy nimi. Widoczna jest także obecność skośnych rozkładów i duża liczba punktów skupionych wzdłuż konkretnych wartości (np. w przypadku cech binarnych **DIVERTED**, **CANCELLED**).



Rysunek 9: Opóźnienia linii

- **Virgin America, JetBlue** oraz **Alaska Airlines** cechują się wąskim zakresem opóźnień – większość lotów miała opóźnienia poniżej 5 minut.
- **Spirit Air Lines** i **Frontier Airlines** odnotowały znacznie większą zmienność, z wieloma przypadkami bardzo długich opóźnień (powyżej 10–15 godzin), co wskazuje na problemy operacyjne i kumulację opóźnień w sieci przewoźnika.
- Średnie i typowe wartości dla **American Airlines, Delta** i **United** zawierają się w przedziale 0–10 minut, ale wykazują pojedyncze ekstremalne przypadki (ponad 20–30 godzin).



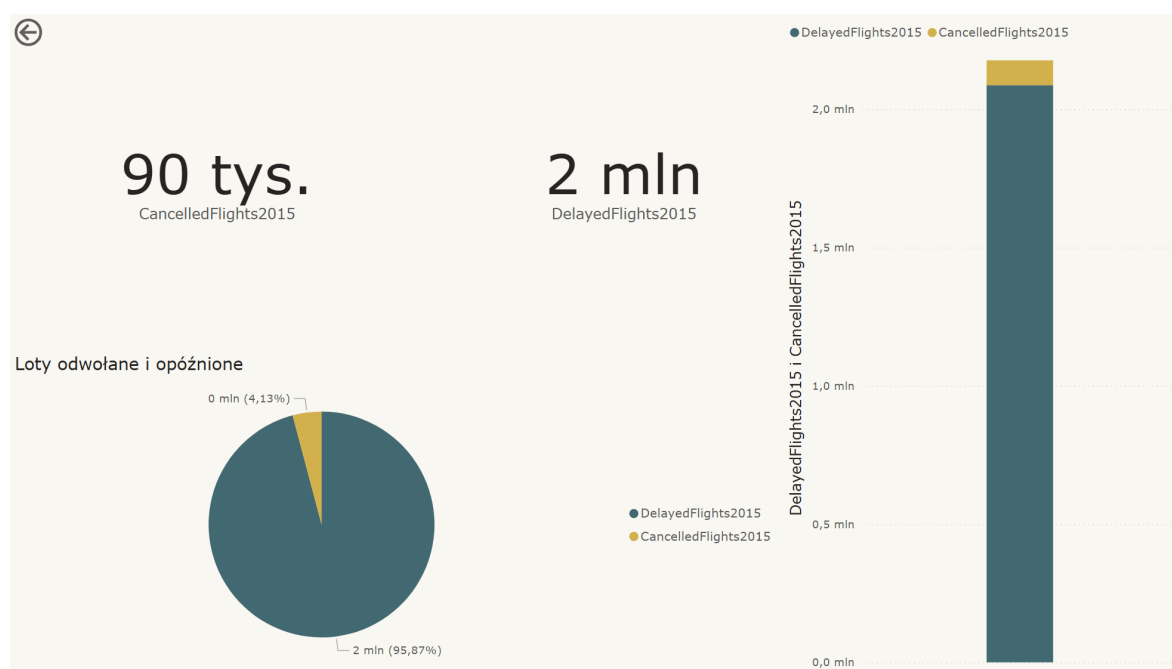
Rysunek 10: Lokalizacja lotnisk (z pakietu 'Maps')

Analiza wizualna odpowiadająca na pytania badawcze wykonana w Power BI, została opisana w kolejnych rozdziałach.

Analiza

Liczba odwołań i opóźnień lotów

Na potrzeby analizy i wizualizacji danych dotyczących opóźnień i odwołań lotów w USA w programie Power BI opracowano zestaw dashboardów przedstawiających różne aspekty funkcjonowania lotnictwa cywilnego. Pierwszy z nich prezentuje ogólne statystyki wykonanych, opóźnionych i odwołanych lotów w skali całego kraju, umożliwiając szybki przegląd skali problemu oraz identyfikację głównych kategorii zakłóceń w ruchu lotniczym.



Rysunek 11: Liczba odwołanych i opóźnionych lotów

Na dashboardzie przedstawiono dwa kluczowe wskaźniki: liczbę odwołanych oraz liczbę opóźnionych lotów w roku 2015. Wskaźniki KPI zostały zaprezentowane w formie dużych, czytelnych wartości liczbowych, ułatwiających szybki wgląd w skalę zjawiska.

W 2015 roku odwołano około **90 tysięcy** lotów, natomiast **ponad 2 miliony** połączeń zostało opóźnionych. Dzięki zastosowaniu filtrów użytkownik ma możliwość zawężenia wyników do konkretnego przewoźnika, lotniska lub przedziału czasowego. Pozwala to na precyzyjną analizę zakłóceń w różnych kontekstach – zarówno przestrzennych, jak i operacyjnych.

Wartości te wskazują, że co trzeci lot w analizowanym roku był zakłócony, co podkreśla skalę problemu punktualności w amerykańskim lotnictwie cywilnym.

Top 10 lotnisk według liczby opóźnień i odwołań

Na kolejnym dashboardzie zaprezentowano dwa wykresy słupkowe obrazujące dziesięć lotnisk z największą liczbą opóźnień (górny wykres) oraz odwołań (dolny wykres).

W zestawieniu opóźnień dominują największe amerykańskie huby lotnicze: *Hartsfield-Jackson Atlanta International Airport* oraz *Chicago O'Hare International Airport*, z ponad 110 tys. opóźnionych lotów każdy. Na kolejnych miejscach znalazły się *Dallas/Fort Worth*, *Denver International* oraz *Los Angeles International*, każdorazowo notując od 60 tys. do 100 tys. opóźnień.

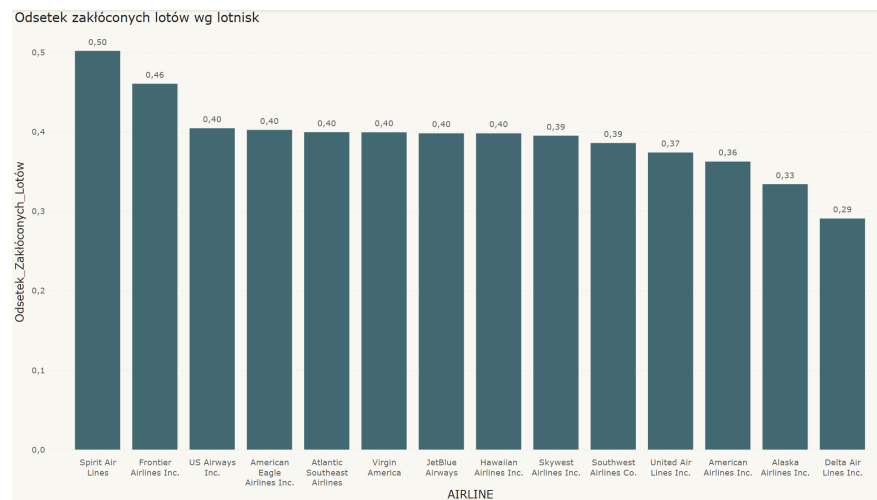


Rysunek 12: Top 10 Opóźnień i odwołań wg lotnisk

Ranking odwołań otwiera ponownie Chicago O'Hare (około 9 tys. anulowanych lotów), a tuż za nim plasują się *Dallas/Fort Worth* oraz *LaGuardia Airport (Marine Air Terminal)* z ponad 4,5 tys. odwołań. W pierwszej dziesiątce znajdują się także Newark Liberty, Logan i Hartsfield-Jackson, co pokazuje, że największe porty lotnicze borykają się zarówno z największą liczbą opóźnień, jak i odwołań.

Najczęściej zakłócone linie

Trzeci dashboard przedstawia zestawienie linii lotniczych według odsetka zakłóconych lotów, czyli sumy lotów opóźnionych i odwołanych względem całkowitej liczby połączeń danego przewoźnika. Wskaźnik ten został zaprezentowany w formie poziomego wykresu słupkowego, ułatwiającego porównanie niezawodności operacyjnej przewoźników.

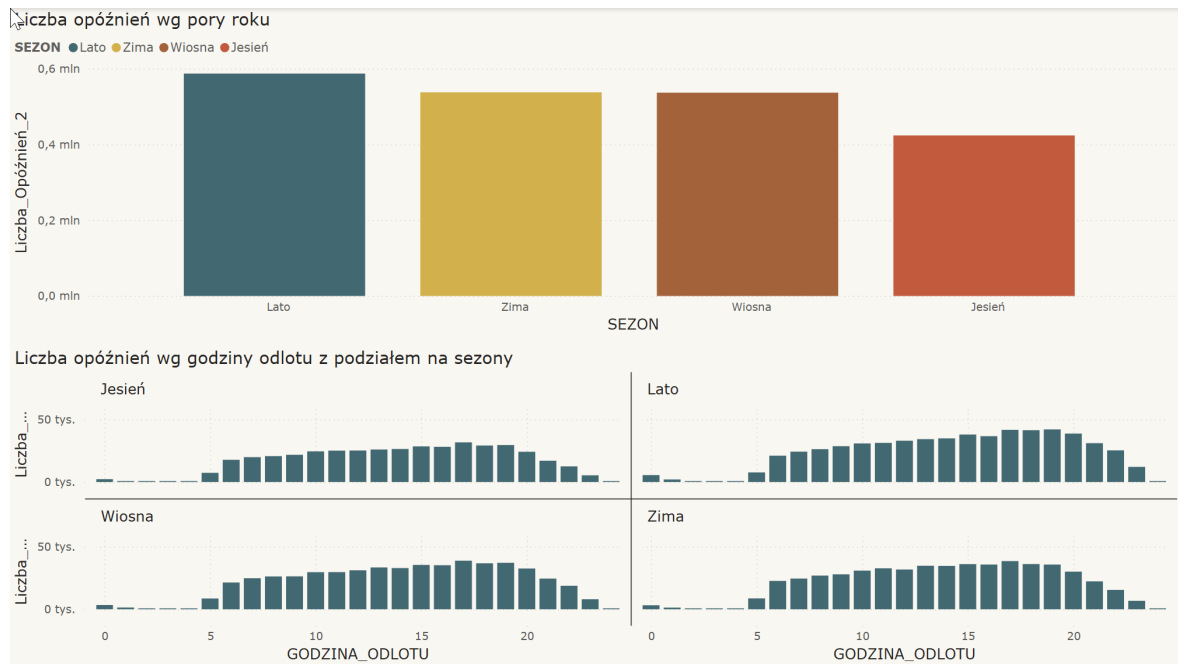


Rysunek 13: Odsetek zakłóconych linii

Najwyższy odsetek zakłóceń występował w przypadku linii Spirit Air Lines, gdzie ponad 50% lotów było opóźnionych lub odwołanych. Na kolejnych pozycjach znalazły się Frontier Airlines Inc. oraz US Airways Inc., również przekraczające próg 40%. Z drugiej strony, Delta Air Lines Inc. oraz Alaska Airlines Inc. osiągnęły najniższe wartości wskaźnika, odpowiednio poniżej 30%, co świadczy o ich wysokiej punktualności.

Wizualizacja pozwala na ocenę ogólnej jakości obsługi pasażerów przez poszczególne przewoźników i może stanowić istotny punkt odniesienia przy wyborze linii lotniczych przez klientów.

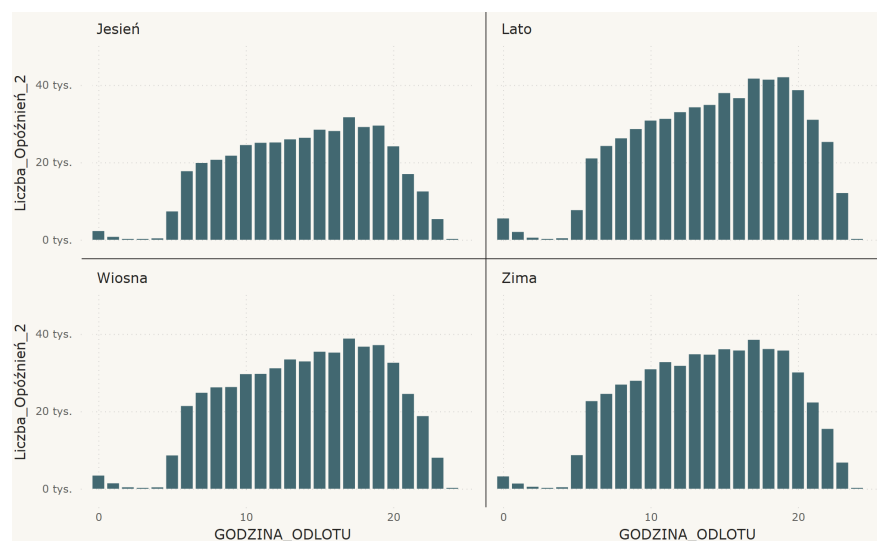
Najczęściej występujące opóźnienia wg pory roku i dnia



Rysunek 14: Dashboard z opóźnieniami wg pory roku i godzinami odlotów

Czwarty dashboard przedstawia rozkład opóźnień lotów w zależności od pory roku oraz godziny odlotu. W górnej części wizualizacji zaprezentowano liczbę opóźnień z podziałem na sezony kalendarzowe, natomiast w dolnej – ich rozkład godzinowy.

Największa liczba opóźnień występowała w okresie letnim, co jest związane z sezonem wakacyjnym i dużym natężeniem ruchu pasażerskiego. Lato wyprzedza pod tym względem zimę oraz wiosnę, natomiast najmniej opóźnień zaobserwowano jesienią.



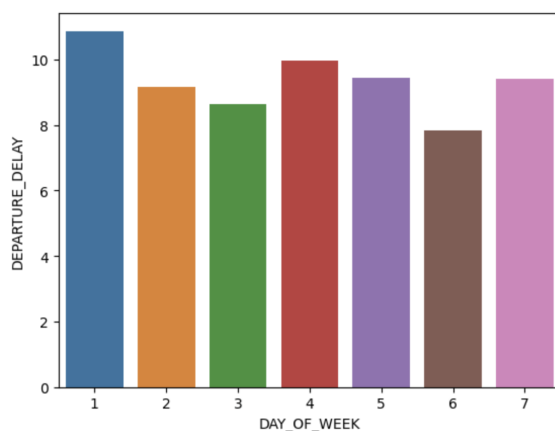
Rysunek 15: Liczba opóźnień wg godziny odlotu z podziałem na sezony

Na wykresie w podziale na cztery sezony widać wspólny, wyraźny trend: liczba

opóźnionych lotów jest znikoma w godzinach nocnych, a od około 6:00 rośnie stopniowo, osiągając szczyt w godzinach popołudniowych.

- **Lato:** Kulminacja opóźnień przypada na przedział 16:00–18:00, kiedy to liczba opóźnionych odlotów przekracza 40 tys.
- **Wiosna i zima:** Maksima obserwowane są nieco wcześniej (około 15:00–17:00), ale wartości pozostają niższe niż latem (35–38 tys.).
- **Jesień:** Przy mniejszym natężeniu ruchu wakacyjnego najwyższa liczba opóźnień sięga około 30 tys., a wzrost do maksimum jest bardziej łagodny.

Taki rozkład wskazuje na silną sezonowość ruchu lotniczego oraz kumulację problemów operacyjnych w godzinach największego obciążenia, zwłaszcza w miesiącach letnich.



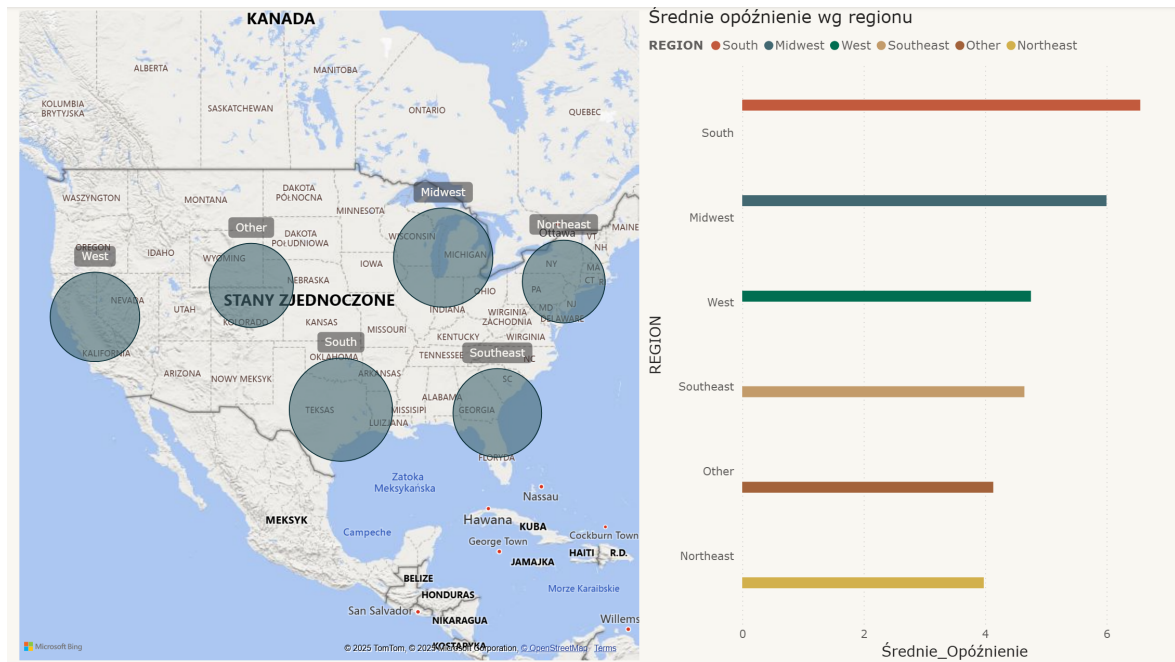
Rysunek 16: Opóźnienia wg dnia tygodnia

Na wykresie słupkowym przedstawiono średni czas opóźnienia odlotu (w minutach) w kolejnych dniach tygodnia (1 – poniedziałek, ..., 7 – niedziela).

Z analizy wynika, że:

- Największe średnie opóźnienie występuje w poniedziałek – przekracza 10 minut, co może być efektem kumulacji problemów operacyjnych po weekendzie.
- W czwartek również obserwujemy wysoki poziom opóźnień (około 10 minut), co wskazuje na silny ruch lotniczy w połowie tygodnia.
- Najniższe opóźnienia występują w sobotę (około 8 minut), co może wynikać z mniejszego natężenia lotów rekreacyjnych.
- Pozostałe dni (wtorek, środa, piątek i niedziela) charakteryzują się średnimi opóźnieniami mieszczącymi się w przedziale od 8,5 do 9,5 minuty.

Różnice pomiędzy regionami



Rysunek 17: Opóźnienia wg regionu

Piąty dashboard umożliwia analizę przestrzenną zakłóceń w ruchu lotniczym, z uwzględnieniem różnic pomiędzy poszczególnymi regionami USA.

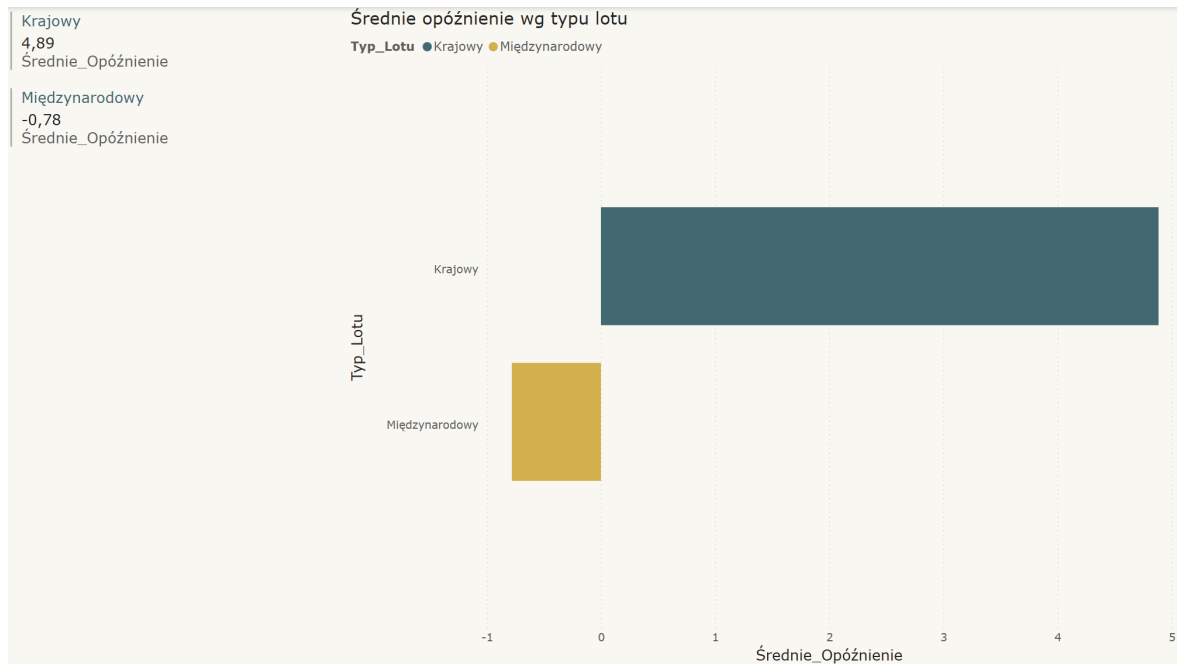
Regiony zostały utworzone poprzez przypisanie stanów do jednej z sześciu grup: *Northeast*, *Southeast*, *South*, *Midwest*, *West* oraz *Other*.

- Wykres kolumnowy po prawej stronie prezentuje **średnie opóźnienie przylotu (ARRIVAL_DELAY)** w minutach dla każdego regionu.
- Mapa po lewej stronie wizualizuje zasięg geograficzny każdego regionu oraz średnie opóźnienie, dzięki rozmiarom kółek reprezentujących zsumowane dane.

Wyniki wskazują, że największe średnie opóźnienia występowały w regionie **South**, a najniższe w regionie **Northeast**. Różnice te mogą wynikać z lokalnych warunków pogodowych, zatłoczenia lotnisk lub charakterystyki operacyjnej linii lotniczych w danym regionie.

Analiza ta pozwala na identyfikację regionów wymagających usprawnień w zakresie zarządzania ruchem lotniczym oraz infrastruktury portów lotniczych.

Połączenia krajowe i międzynarodowe



Rysunek 18: Porównanie opóźnień połączeń krajowych i międzynarodowych

Szósty dashboard przedstawia porównanie średniego czasu opóźnień dla dwóch typów połączeń lotniczych: krajowych i międzynarodowych. Wartości zostały zaprezentowane na wykresie słupkowym, ukazującym wyraźny kontrast pomiędzy tymi kategoriami.

Z analizy wynika, że średnie opóźnienie dla lotów krajowych wynosiło około 4,9 minuty, natomiast dla lotów międzynarodowych średnia była nieznacznie ujemna, co wskazuje na przypadki wcześniejszych odlotów lub brak istotnych opóźnień w tej kategorii. Może to również wynikać z różnic w metodzie rejestracji opóźnień lub specyfiki danych.

Wyniki te sugerują, że loty międzynarodowe, choć rzadziej wykonywane w porównaniu do krajowych, cechują się większą punktualnością. Może to wynikać z priorytetowego traktowania ich w slotach czasowych lub innej strukturze operacyjnej przewoźników.

Wnioski

Na podstawie analizy zbioru „2015 Flight Delays and Cancellations” można sformułować następujące wnioski:

- **Skala zakłóceń była znacząca.** W 2015 roku odnotowano około 2 mln opóźnionych lotów oraz 90 tys. odwołań, co oznacza, że ponad 1 na 3 połączenia zostało zakłócone.
- **Najbardziej obciążone lotniska:** Hartsfield–Jackson Atlanta i Chicago O’Hare wygenerowały najwięcej opóźnień (ponad 110 tys. każde), a port O’Hare dominował również w rankingu odwołań (ok. 9 tys.).
- **Przewoźnicy o najwyższym odsetku zakłóceń:** Spirit Air Lines (50 %) i Frontier Airlines (46 %) miały najwięcej opóźnień i odwołań w stosunku do liczby rejsów. Najniższy wskaźnik odnotowała Delta Air Lines (29 %), następnie Alaska Airlines.
- **Sezonowość i dobowy rozkład opóźnień:** Najwięcej opóźnień przypadało na okres letni, a dobowe maksimum w godzinach 15:00–18:00. Wiosną i zimą szczyty przesunęły się na 15:00–17:00, jesienią liczba opóźnień była najniższa i wzrost do maksimum przebiegał łagodniej.
- **Różnice regionalne:** Najwięcej zakłóceń odnotowano w Kalifornii (i 250 tys.), Illinois i na Florydzie. Jednocześnie najwyższe średnie czasy opóźnień występowały w małych stanach, takich jak Delaware (24 min) i Samoa Amerykańskie (21,5 min).
- **Typ lotu:** Loty krajowe miały średnie opóźnienie blisko 5 min, natomiast loty międzynarodowe wykazały średnio zerowe lub minimalnie ujemne wartości, co świadczy o większej punktualności tego segmentu.
- **Rozkład w tygodniu:** Największe średnie opóźnienia występowały w poniedziałek i czwartek (około 10 min), najniższe w sobotę (8 min), co wskazuje na kumulację problemów operacyjnych na początku i w połowie tygodnia.

Ocena postawionych hipotez

- **Sezonowość opóźnień:** potwierdzona — lato wykazało najwyższy poziom opóźnień.
- **Loty poranne mają największy odsetek opóźnień:** obalona — szczyt opóźnień występuje popołudniu (15:00–18:00), nie rano.
- **Duże lotniska międzynarodowe notują więcej opóźnień:** potwierdzona — największe porty (Atlanta, O'Hare, Dallas/Fort Worth) dominują w rankingu opóźnień.
- **Linie budżetowe mają wyższy odsetek odwołań:** potwierdzona — przewoźnicy ultra low-cost (Spirit, Frontier) osiągnęli najwyższe wskaźniki zakłóceń.
- **Loty wieczorne są bardziej podatne na opóźnienia:** obalona — największe opóźnienia kumulują się w godzinach popołudniowych, a wieczorne odloty mają niższe wartości.

Zakłócenia w ruchu lotniczym są zatem zjawiskiem wielowymiarowym — zależą od pory roku, dnia tygodnia, lokalizacji, typu lotu i praktyk operacyjnych przewoźników. Wyniki tej analizy mogą posłużyć do optymalizacji grafików rejsów, lepszego planowania zasobów oraz doskonalenia systemów prognozowania warunków operacyjnych.

Bibliografia

- [1] U.S. Department of Transportation, Bureau of Transportation Statistics. *2015 Flight Delays and Cancellations*. Kaggle Dataset, 2015. Dostęp: czerwiec 2025, <https://www.kaggle.com/datasets/usdot/flight-delays>