

Bartosz Szyca  
Jolanta Kunicka

# Predykcja opóźnień samolotów

---

P r e s e n t a t i o n



# 0 projekcie

Naszym głównym celem jest opracowanie modelu predykcyjnego, który pozwoli nam zrozumieć i przewidzieć, **czy** samolot będzie opóźniony czy nie. Jest to kluczowy krok w minimalizowaniu opóźnień w ruchu lotniczym i poprawie jakości usług oferowanych przez linie lotnicze.





# Źródło danych:

Dane, które wykorzystujemy do analizy i budowy modelu predykcyjnego w naszym projekcie pochodzą z platformy Kaggle, która jest znanym źródłem zbiorów danych i projektów z różnych dziedzin. Konkretnie, korzystamy z zestawu danych udostępnionego na Kaggle pod adresem [link](#).

Zestaw danych zawiera szeroki zakres informacji związanych z opóźnieniami lotów w Stanach Zjednoczonych. Dane te zostały zebrane przez amerykański Departament Transportu (USDOT) oraz inne agencje rządowe, co sprawia, że są one źródłem godnym zaufania i kompetentnym w dziedzinie ruchu lotniczego.

Dzięki dostępowi do tych danych, jesteśmy w stanie przeprowadzić dogłębną analizę opóźnień samolotów oraz opracować model predykcyjny, który pomoże nam zrozumieć i przewidzieć, czy samolot będzie opóźniony czy nie. Nasza praca opiera się na dokładnych i wiarygodnych źródłach danych, co stanowi fundament naszego projektu. W dalszej części prezentacji przedstawimy strukturę i zawartość tych danych oraz planowane kroki w analizie i modelowaniu.



# Struktura danych:

- 01 Data i godzina odlotu
- 02 Linia lotnicza
- 03 Numer lotu
- 04 Numer samolotu (tail number)
- 05 Planowana godzina wylotu
- 06 Lotnisko wylotu
- 07 Lotnisko docelowe
- 08 Planowana długość lotu
- 09 Planowana godzina przylotu
- 10 Długość trasy

# Typy danych:

01

## Dane liczbowe

Planowana długość lotu  
Planowana długość trasy

02

## Dane katagoryczne

Linie lotnicze  
Kody lotnisk wylotu i przylotu  
Numery lotów

03

## Dane czasowe

Planowana data, godzina wylotu, przylotu

[ ] X

	DAY_sin	DAY_cos	DAY_OF_WEEK_sin	DAY_OF_WEEK_cos	MONTH_sin	MONTH_cos	ID_AIRLINE	FLIGHT_NUMBER	ID_TAIL_NUMBER	SCHEDULED_DEPARTURE_sin	SCHEDULED_DEPARTURE_cos	ID_ORIGIN_AIRPORT	ID_DESTINATION_AIRPORT	SCHEDULED_TIME	SCHEDULED_ARRIVAL_sin	SCHEDULED_ARRIVAL_cos	DISTANCE
0	2.012985e-01	0.97953	-0.433884	-0.900969	5.000000e-01	0.866025	1	98	1622	0.021815	0.999762	17.0	277.0	205.0	0.923880	0.382683	1448
1	2.012985e-01	0.97953	-0.433884	-0.900969	5.000000e-01	0.866025	0	2336	1556	0.043619	0.999048	176.0	235.0	280.0	0.887011	-0.461749	2330
2	2.012985e-01	0.97953	-0.433884	-0.900969	5.000000e-01	0.866025	11	840	421	0.087156	0.996195	278.0	66.0	286.0	0.852640	-0.522499	2296
3	2.012985e-01	0.97953	-0.433884	-0.900969	5.000000e-01	0.866025	0	258	1516	0.087156	0.996195	176.0	203.0	285.0	0.854912	-0.518773	2342
4	2.012985e-01	0.97953	-0.433884	-0.900969	5.000000e-01	0.866025	1	135	2131	0.108867	0.994056	277.0	17.0	235.0	0.766044	0.642788	1448
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5819074	-2.449294e-16	1.00000	-0.433884	-0.900969	-2.449294e-16	1.000000	2	688	2885	-0.004363	0.999990	176.0	39.0	320.0	0.821647	-0.569997	2611
5819075	-2.449294e-16	1.00000	-0.433884	-0.900969	-2.449294e-16	1.000000	2	745	3947	-0.004363	0.999990	166.0	248.0	227.0	0.948324	0.317305	1617
5819076	-2.449294e-16	1.00000	-0.433884	-0.900969	-2.449294e-16	1.000000	2	1503	4416	-0.004363	0.999990	166.0	285.0	221.0	0.939693	0.342020	1598
5819077	-2.449294e-16	1.00000	-0.433884	-0.900969	-2.449294e-16	1.000000	2	333	2132	-0.004363	0.999990	193.0	285.0	161.0	0.819152	0.573576	1189
5819078	-2.449294e-16	1.00000	-0.433884	-0.900969	-2.449294e-16	1.000000	2	839	2169	-0.004363	0.999990	166.0	42.0	221.0	0.939693	0.342020	1576

5714008 rows x 17 columns

y

0 -22.0