Projekt IUM - 2021Z - Zad.3

Aleksander Garlikowski
Cezary Moczulski

Problem biznesowy:

Przewidywanie czasu transportu zakupionego produktu.

Zadanie modelowania:

Regresja czasu dostawy na podstawie dnia tygodnia zamówienia, miasta docelowego oraz firmy transportowej obsługującej zamówienie.

Biznesowe kryterium sukcesu:

Przekazywanie klientom przewidywań na temat zakresu czasu dostawy, dokładniejszych i bardziej wiarygodnych niż rozsądne przewidywanie, że produkt zostanie dostarczony w zakresie od 24 do 72 godzin od momentu zakupu (na podstawie dostępnych danych sprawdza się ono w ~79% przypadków).

Analityczne kryterium sukcesu:

Uzyskanie przewidywanego zakresu czasu węższego niż 48 godzin, przy celności wyższej niż 90% (żeby można było uznać przewidywanie za znacznie bardziej wiarygodne i zostawić margines błędu dla wdrożenia modelu), na posiadanych danych.

90% celności powinno być wynikiem osiągalnym, ponieważ nawet przewidywanie polegające na podawaniu średniej dla każdej z kombinacji miasta, firmy kurierskiej i dnia tygodnia osiągają ponad 90% skuteczności z oknem czasowym o szerokości 48 godzin.

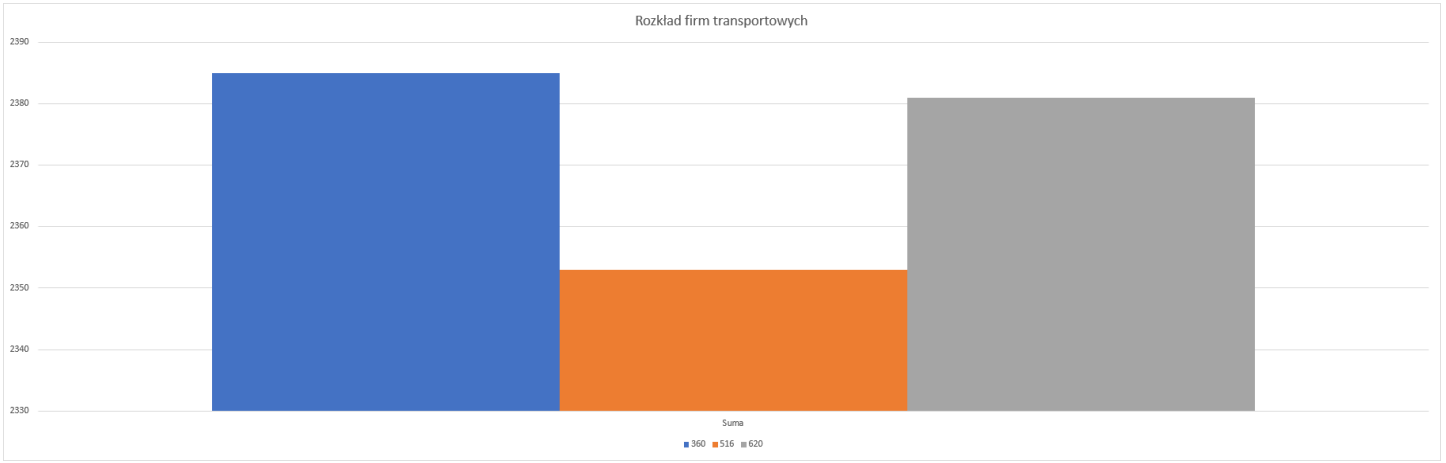
Analiza danych:

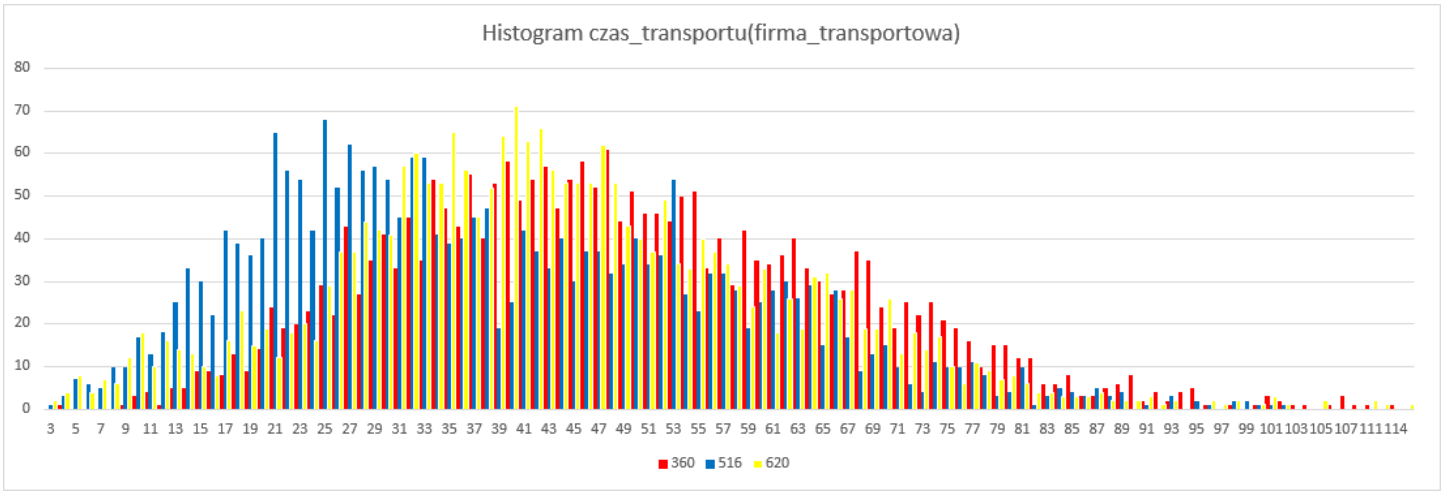
Wybrane dane:

- Firma kurierska
- Miasto docelowe
- Dzień tygodnia złożenia zamówienia

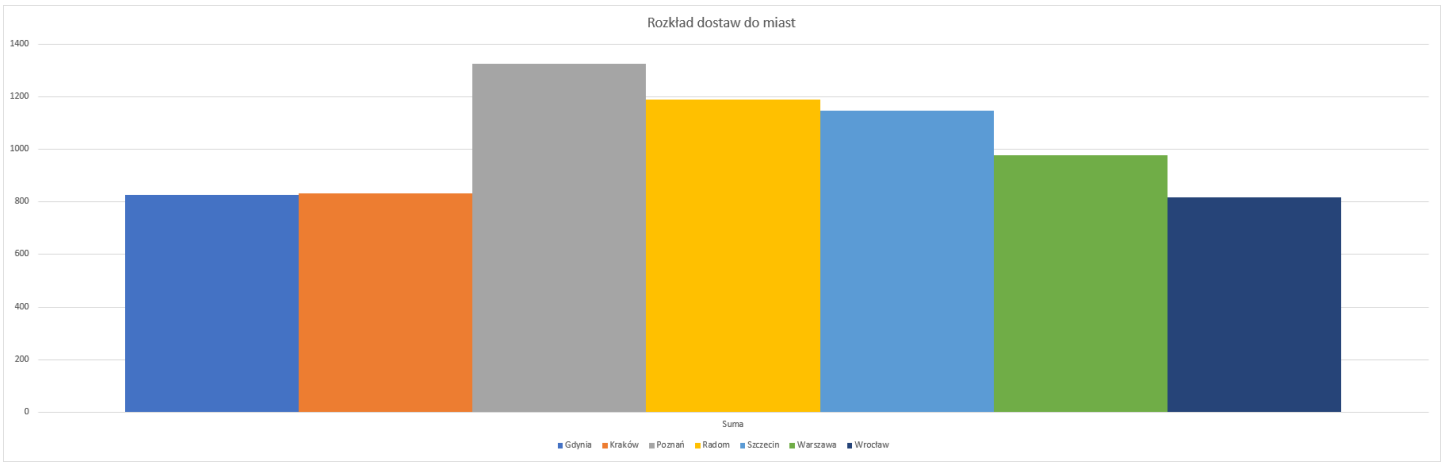
Poniżej znajdują się wykresy rozkładu wystąpień atrybutów w danych i rozkłady czasu transportu (zmiennej celu) od tych atrybutów.

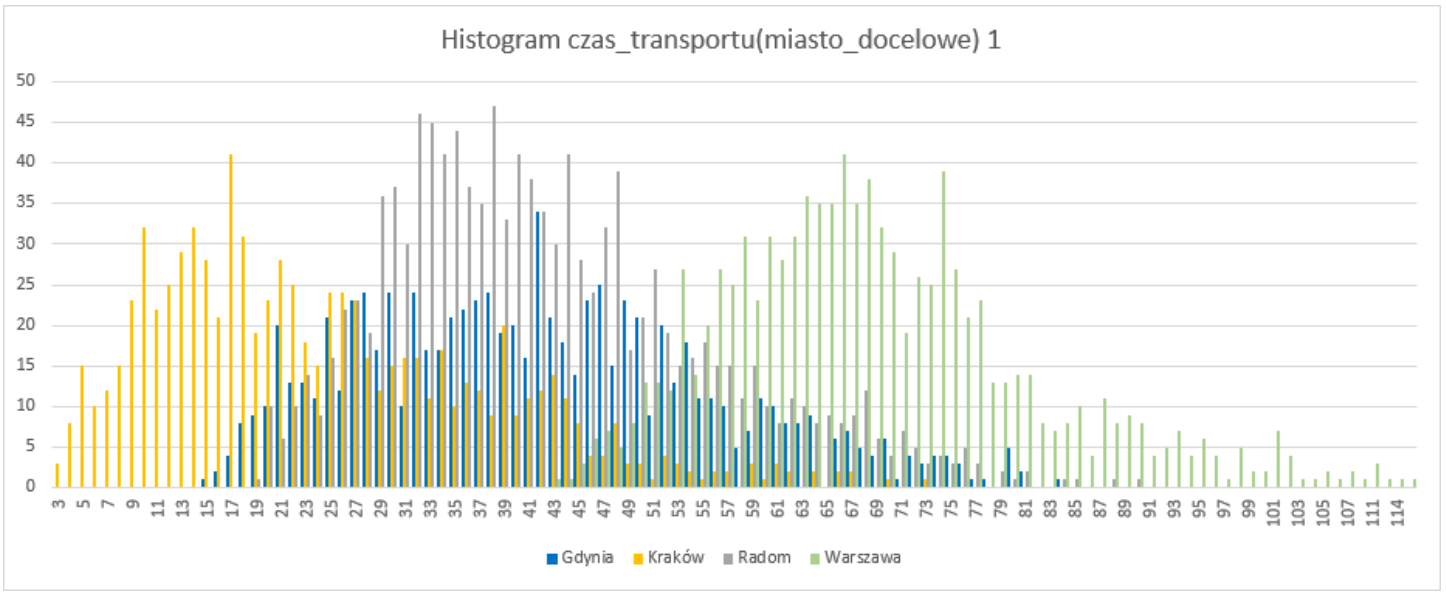
Firma kurierska:

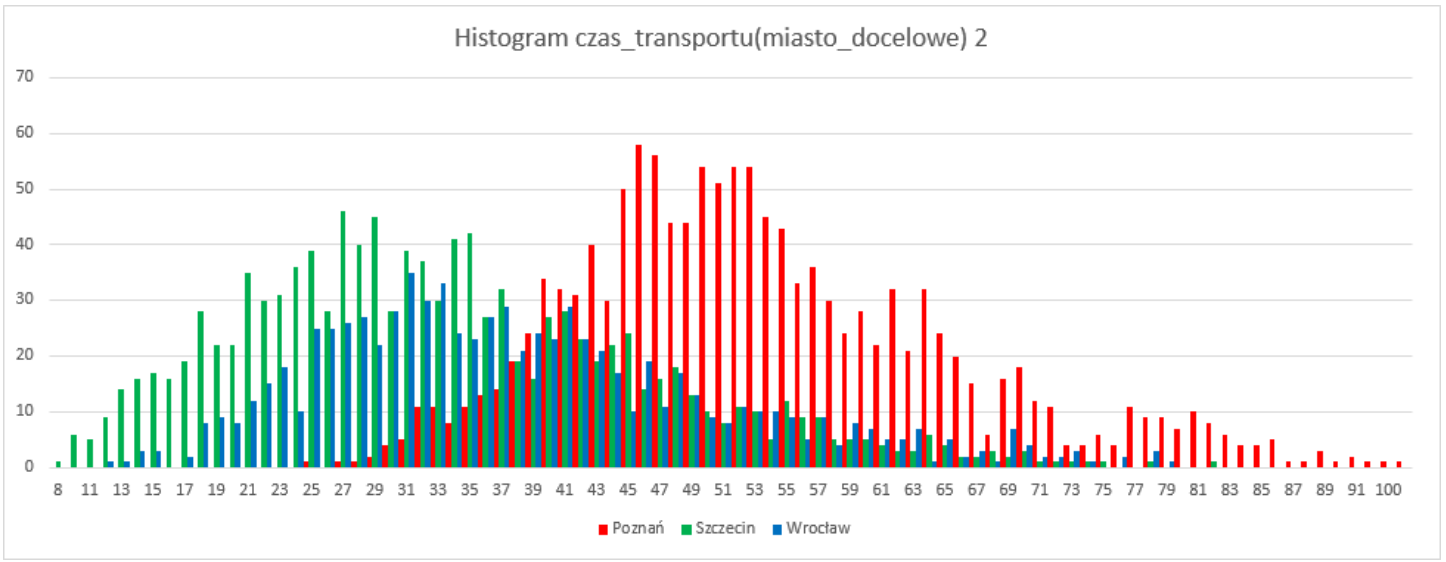




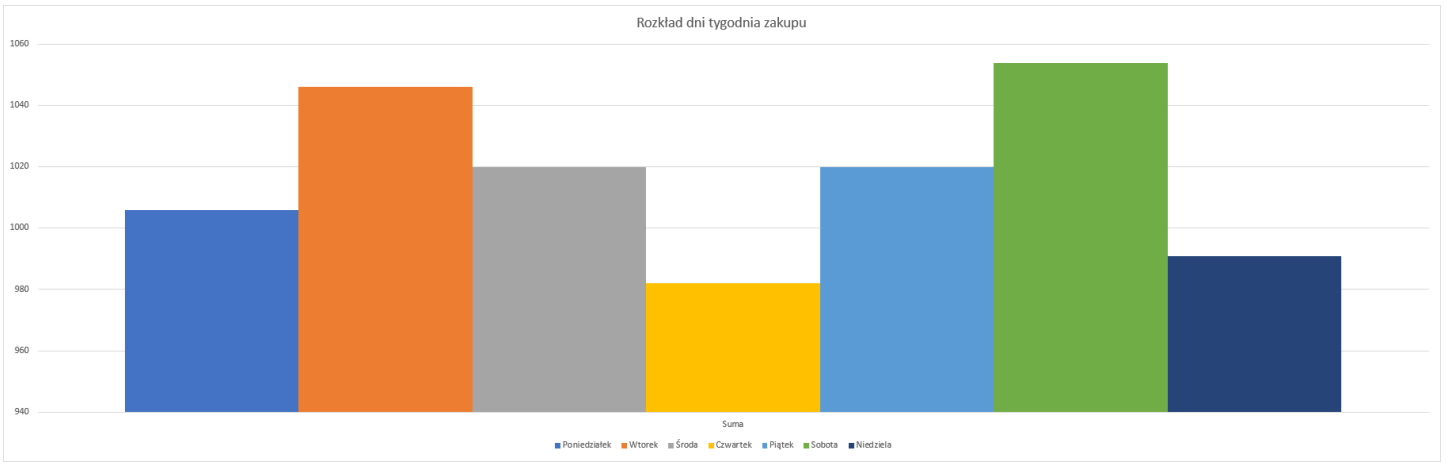
Miasto:

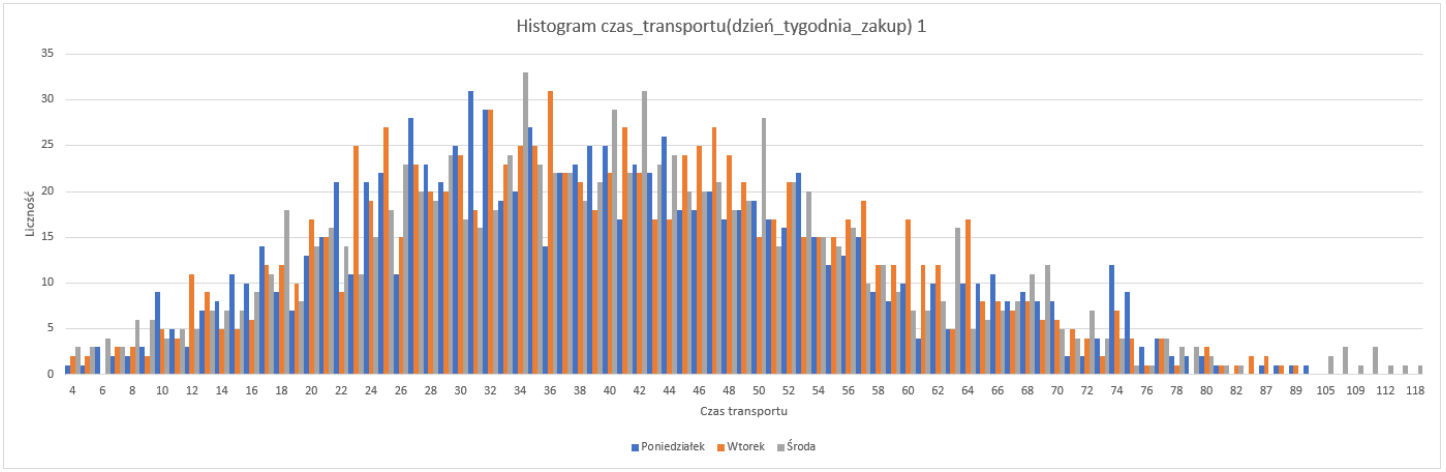

Rozkład dostaw do miast
1400
1200
1000
800
600
400
200
0
Suma
Gdynia
Kraków
Poznań
Radom
Szczecin
Warszawa
Wrocław

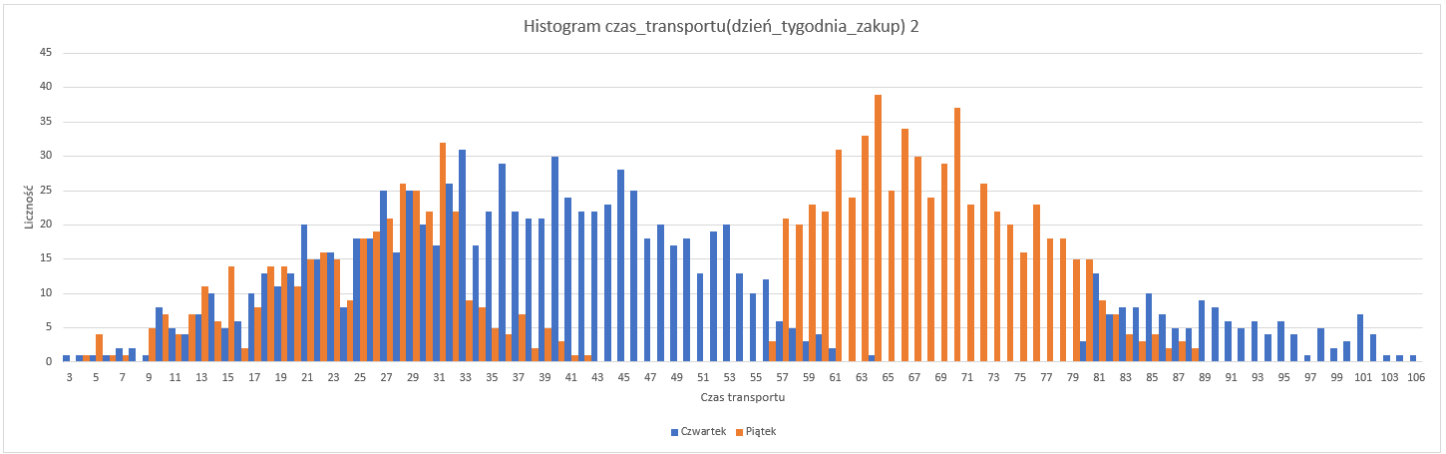

Histogram czas_transportu(miasto_docelowe) 1
50
45
40
35
30
25
20
15
10
5
0
3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89 91 93 95 97 99 101 103 105 107 111 114
Gdynia
Kraków
Radom
Warszawa

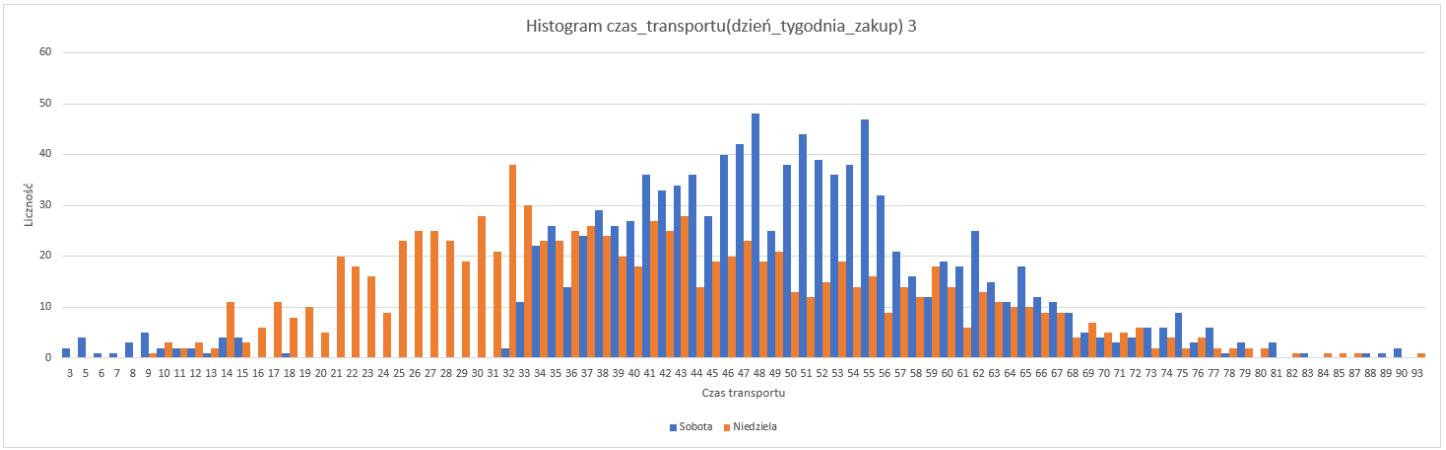

Histogram czas_transportu(miasto_docelowe) 2
70
60
50
40
30
20
10
0
8 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89 91 100
Poznań
Szczecin
Wrocław

Dzień tygodnia:









Odrzucone dane:

Odrzucone na podstawie wiedzy dziedzinowej:

- Imię i nazwisko klienta
- Nazwa produktu
- Cena produktu
- Przecena

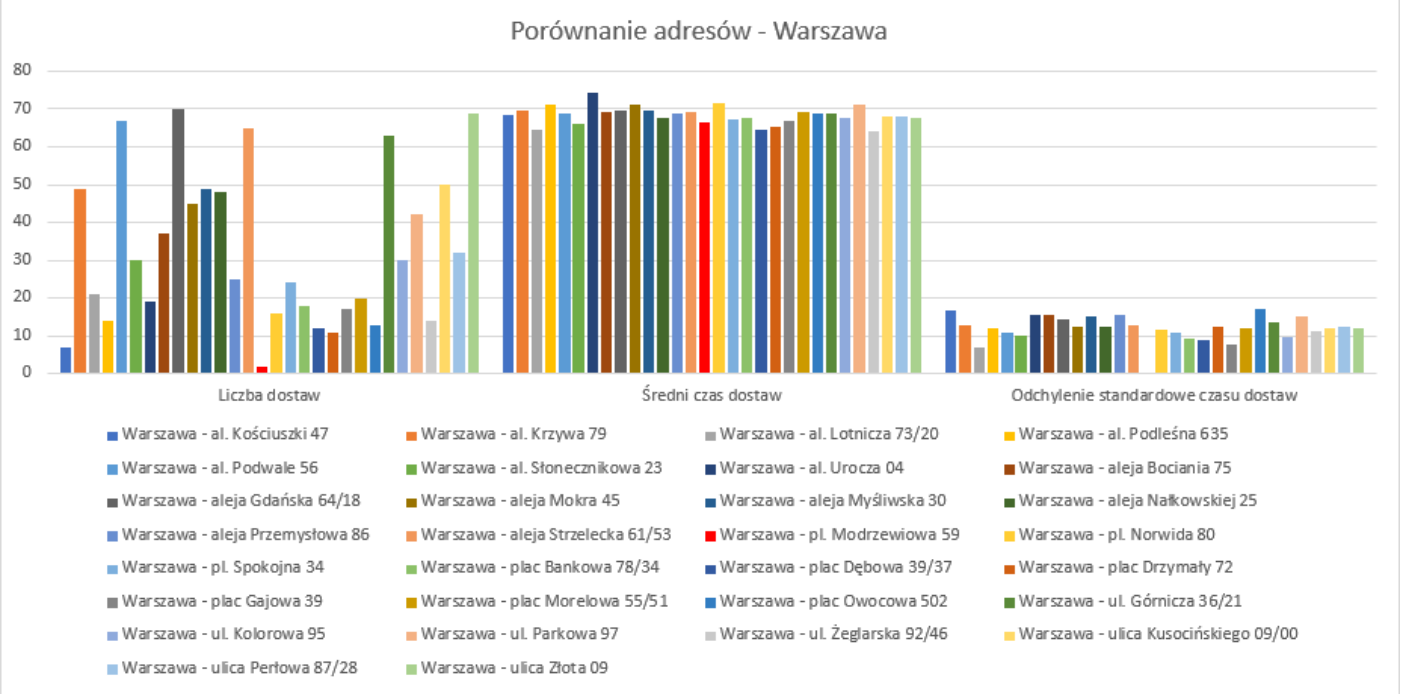
Odrzucone, ponieważ nie niosą znaczącej informacji o zmiennej celu:

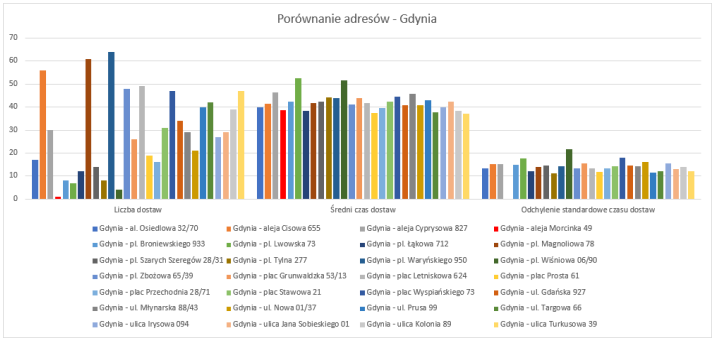
- Nazwy ulic (adresy)
- Kategoria produktu (gabaryty najwidoczniej nie mają znaczenia)
- Miesiąc
- Dokładna data (np. święta)
- Godzina

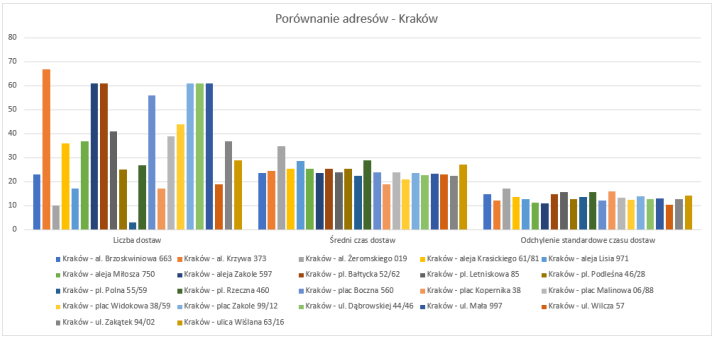
Ponieważ nie mamy danych pochodzących z innych lat, nie jesteśmy w stanie stwierdzić czy rok niesie ze sobą istotne informacje. Z daty uwzględniamy więc jedynie dzień tygodnia.

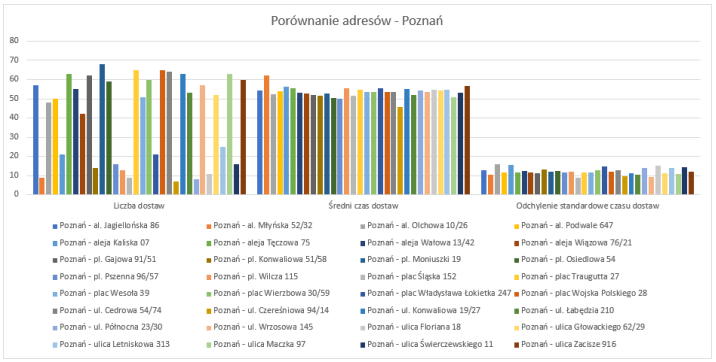
Po przejrzeniu danych stwierdziliśmy, że rozkład czasu transportu (zmiennej celu) dla konkretnej wartości każdego z odrzuconych atrybutów jest bardzo podobny do rozkładów dla jego pozostałych wartości.
Poniżej są wykresy średnich i odchyleń standardowych odrzuconych atrybutów.

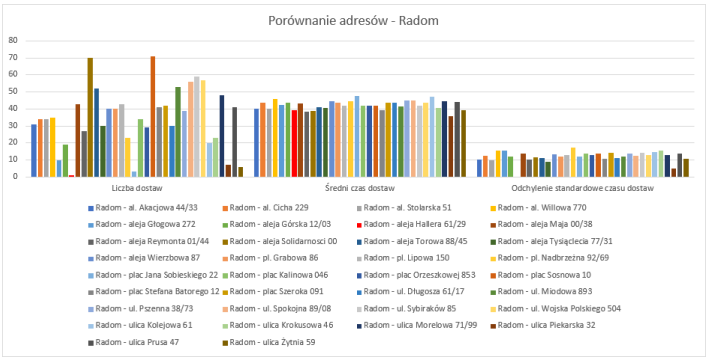
Adresy w ramach miasta:

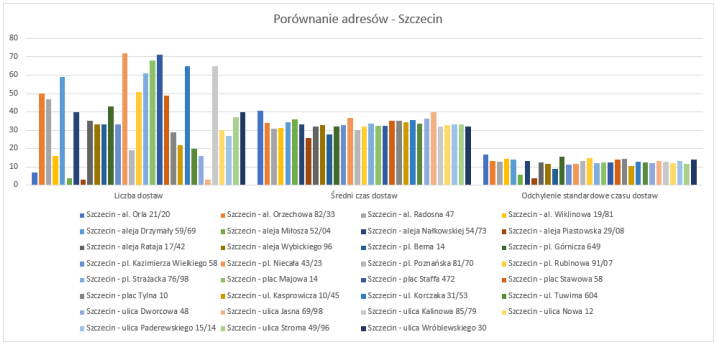


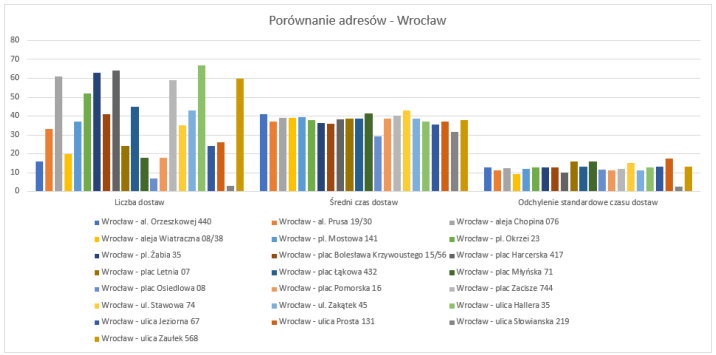

Porównanie adresów - Gdynia
Liczba dostaw
Średni czas dostaw
Odchylenie standardowe czasu dostaw
Gdynia - al. Osiedlowa 32/70
Gdynia - aleja Cisowa 655
Gdynia - aleja Cyprysowa 827
Gdynia - aleja Morcinka 49
Gdynia - pl. Broniewskiego 933
Gdynia - pl. Lwowska 73
Gdynia - pl. Łąkowa 712
Gdynia - pl. Magnoliowa 78
Gdynia - pl. Szarych Szeregów 28/31
Gdynia - pl. Tylna 277
Gdynia - pl. Waryńskiego 950
Gdynia - pl. Wiśniowa 06/90
Gdynia - pl. Zbożowa 65/39
Gdynia - plac Grunwaldzka 53/13
Gdynia - plac Letniskowa 624
Gdynia - plac Prosta 61
Gdynia - plac Przechodnia 28/71
Gdynia - plac Stawowa 21
Gdynia - plac Wyspiańskiego 73
Gdynia - ul. Gdańska 927
Gdynia - ul. Młynarska 88/43
Gdynia - ul. Nowa 01/37
Gdynia - ul. Prusa 99
Gdynia - ul. Targowa 66
Gdynia - ulica Irysowa 094
Gdynia - ulica Jana Sobieskiego 01
Gdynia - ulica Kolonia 89
Gdynia - ulica Turkusowa 39


Porównanie adresów - Kraków
Liczba dostaw
Średni czas dostaw
Odchylenie standardowe czasu dostaw
Kraków - al. Brzoskwiniowa 663
Kraków - al. Krzywa 373
Kraków - al. Żeromskiego 019
Kraków - aleja Krasickiego 61/81
Kraków - aleja Lisia 971
Kraków - aleja Miłosza 750
Kraków - aleja Zakole 597
Kraków - pl. Bałtycka 52/62
Kraków - pl. Letniskowa 85
Kraków - pl. Podleśna 46/28
Kraków - pl. Polna 55/59
Kraków - pl. Rzeczna 460
Kraków - plac Boczna 560
Kraków - plac Kopernika 38
Kraków - plac Malinowa 06/88
Kraków - plac Widokowa 38/59
Kraków - plac Zakole 99/12
Kraków - ul. Dąbrowskiej 44/46
Kraków - ul. Mała 997
Kraków - ul. Wilcza 57
Kraków - ul. Zakątek 94/02
Kraków - ulica Wiślana 63/16

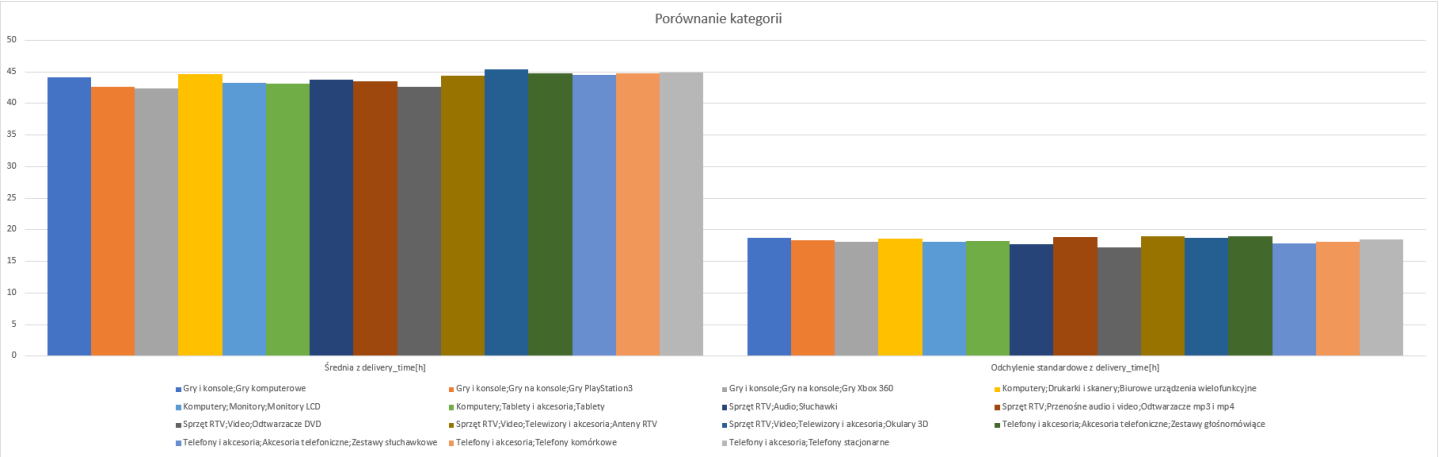

Porównanie adresów - Poznań
Liczba dostaw
Średni czas dostaw
Odchylenie standardowe czasu dostaw
Poznań - al. Jagiellońska 86
Poznań - al. Młyńska 52/32
Poznań - al. Olchowa 10/26
Poznań - al. Podwale 647
Poznań - aleja Kaliska 07
Poznań - aleja Tęczowa 75
Poznań - aleja Wałowa 13/42
Poznań - aleja Wiązowa 76/21
Poznań - pl. Gajowa 91/51
Poznań - pl. Konwaliowa 51/58
Poznań - pl. Moniuszki 19
Poznań - pl. Osiedlowa 54
Poznań - pl. Pszenna 96/57
Poznań - pl. Wilcza 115
Poznań - plac Śląska 152
Poznań - plac Traugutta 27
Poznań - plac Wesoła 39
Poznań - plac Wierzbowa 30/59
Poznań - plac Władysława Łokietka 247
Poznań - plac Wojska Polskiego 28
Poznań - ul. Cedrowa 54/74
Poznań - ul. Czereśniowa 94/14
Poznań - ul. Konwaliowa 19/27
Poznań - ul. Łabędzia 210
Poznań - ul. Północna 23/30
Poznań - ul. Wrzosowa 145
Poznań - ulica Floriana 18
Poznań - ulica Głowackiego 62/29
Poznań - ulica Letniskowa 313
Poznań - ulica Maczka 97
Poznań - ulica Świerczewskiego 11
Poznań - ulica Zacisze 916


Porównanie adresów - Radom
Liczba dostaw
Średni czas dostaw
Odchylenie standardowe czasu dostaw
Radom - al. Akacjowa 44/33
Radom - al. Cicha 229
Radom - al. Stolarska 51
Radom - al. Willowa 770
Radom - aleja Głogowa 272
Radom - aleja Górska 12/03
Radom - aleja Hallera 61/29
Radom - aleja Maja 00/38
Radom - aleja Reymonta 01/44
Radom - aleja Solidarności 00
Radom - aleja Torowa 88/45
Radom - aleja Tysiąclecia 77/31
Radom - aleja Wierzbowa 87
Radom - pl. Grabowa 86
Radom - pl. Lipowa 150
Radom - pl. Nadbrzeżna 92/69
Radom - plac Jana Sobieskiego 22
Radom - plac Kalinowa 046
Radom - plac Orzeszkowej 853
Radom - plac Sosnowa 10
Radom - plac Stefana Batorego 12
Radom - plac Szeroka 091
Radom - ul. Długosza 61/17
Radom - ul. Miodowa 893
Radom - ul. Pszenna 38/73
Radom - ul. Spokojna 89/08
Radom - ul. Sybiraków 85
Radom - ul. Wojska Polskiego 504
Radom - ulica Kolejowa 61
Radom - ulica Krokusowa 46
Radom - ulica Morelowa 71/99
Radom - ulica Piekarska 32
Radom - ulica Prusa 47
Radom - ulica Żytnia 59

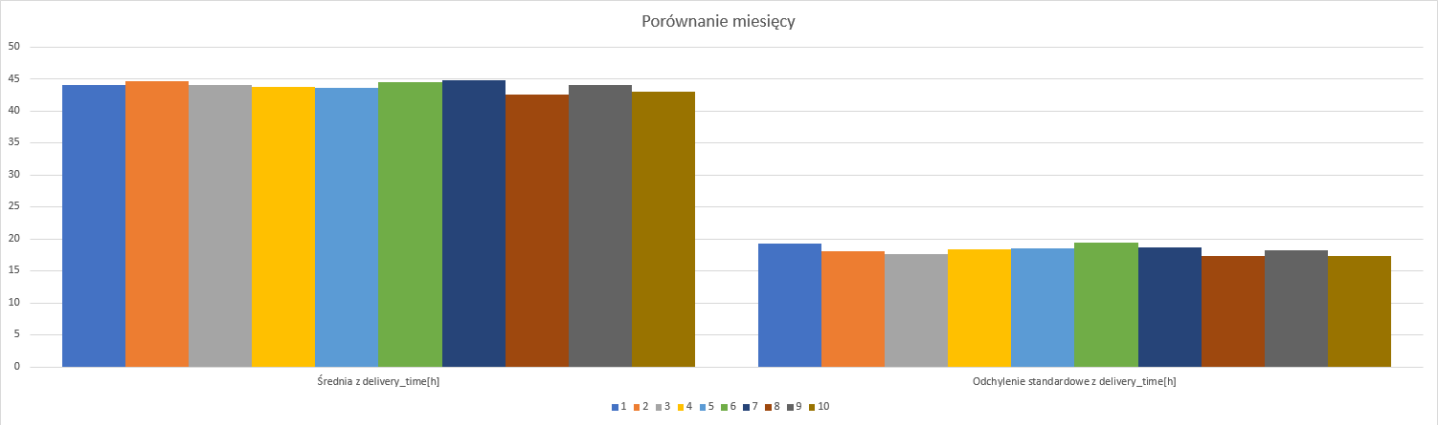

Porównanie adresów - Szczecin
Liczba dostaw
Średni czas dostaw
Odchylenie standardowe czasu dostaw
Szczecin - al. Orla 21/20
Szczecin - al. Orzechowa 82/33
Szczecin - al. Radosna 47
Szczecin - al. Wiklinowa 19/81
Szczecin - aleja Drzymały 59/69
Szczecin - aleja Miłosza 52/04
Szczecin - aleja Natkowskiej 54/73
Szczecin - aleja Piastowska 29/08
Szczecin - aleja Rataja 17/42
Szczecin - aleja Wybickiego 96
Szczecin - pl. Bema 14
Szczecin - pl. Górnicza 649
Szczecin - pl. Kazimierza Wielkiego 58
Szczecin - pl. Niecała 43/23
Szczecin - pl. Poznańska 81/70
Szczecin - pl. Rubinowa 91/07
Szczecin - pl. Strażacka 76/98
Szczecin - plac Majowa 14
Szczecin - plac Staffa 472
Szczecin - plac Stawowa 58
Szczecin - plac Tylna 10
Szczecin - ul. Kasprowicza 10/45
Szczecin - ul. Korczaka 31/53
Szczecin - ul. Tuwima 604
Szczecin - ulica Dworcowa 48
Szczecin - ulica Jasna 69/98
Szczecin - ulica Kalinowa 85/79
Szczecin - ulica Nowa 12
Szczecin - ulica Paderewskiego 15/14
Szczecin - ulica Stroma 49/96
Szczecin - ulica Wróblewskiego 30


Porównanie adresów - Wrocław
Liczba dostaw
Średni czas dostaw
Odchylenie standardowe czasu dostaw
Wrocław - al. Orzeszkowej 440
Wrocław - al. Prusa 19/30
Wrocław - aleja Chopina 076
Wrocław - aleja Wiatraczna 08/38
Wrocław - pl. Mostowa 141
Wrocław - pl. Okrzei 23
Wrocław - pl. Żabia 35
Wrocław - plac Bolesława Krzywoustego 15/56
Wrocław - plac Harcerska 417
Wrocław - plac Letnia 07
Wrocław - plac Łąkowa 432
Wrocław - plac Młyńska 71
Wrocław - plac Osiedlowa 08
Wrocław - plac Pomorska 16
Wrocław - plac Zaciszе 744
Wrocław - ul. Stawowa 74
Wrocław - ul. Zakątek 45
Wrocław - ulica Hallera 35
Wrocław - ulica Jeziorna 67
Wrocław - ulica Prosta 131
Wrocław - ulica Słowiańska 219
Wrocław - ulica Zaułek 568

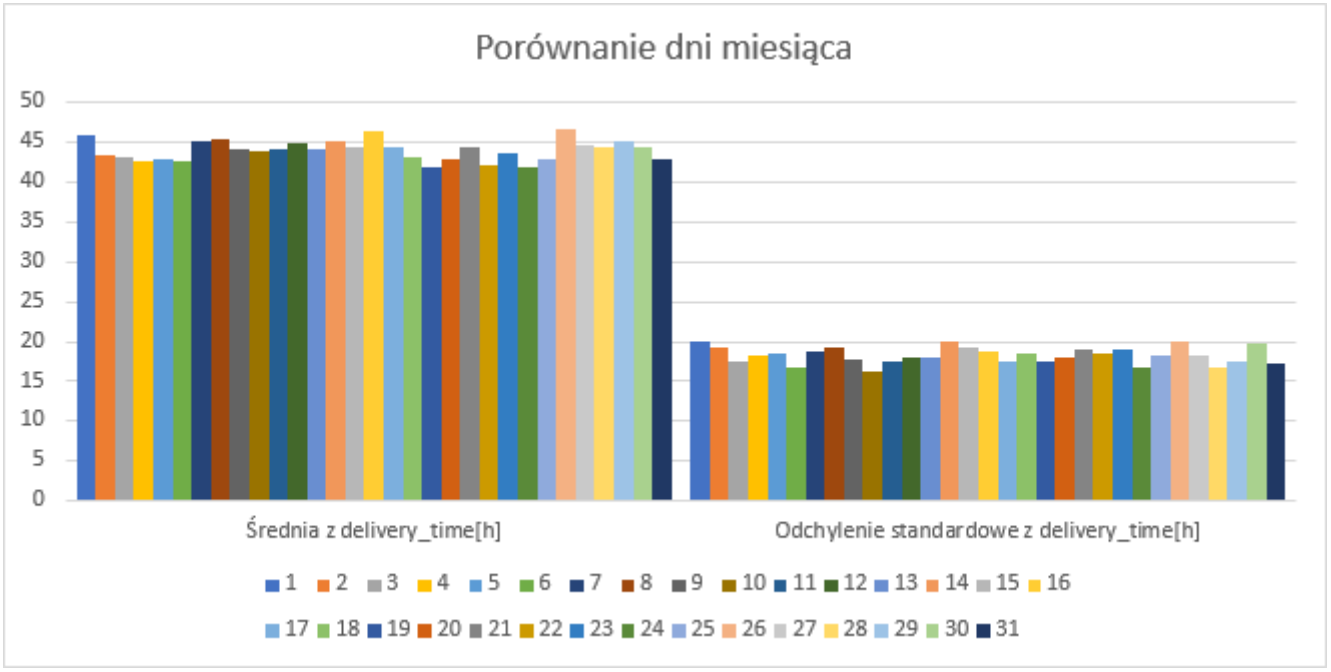
Kategoria produktu:



Miesiąc:



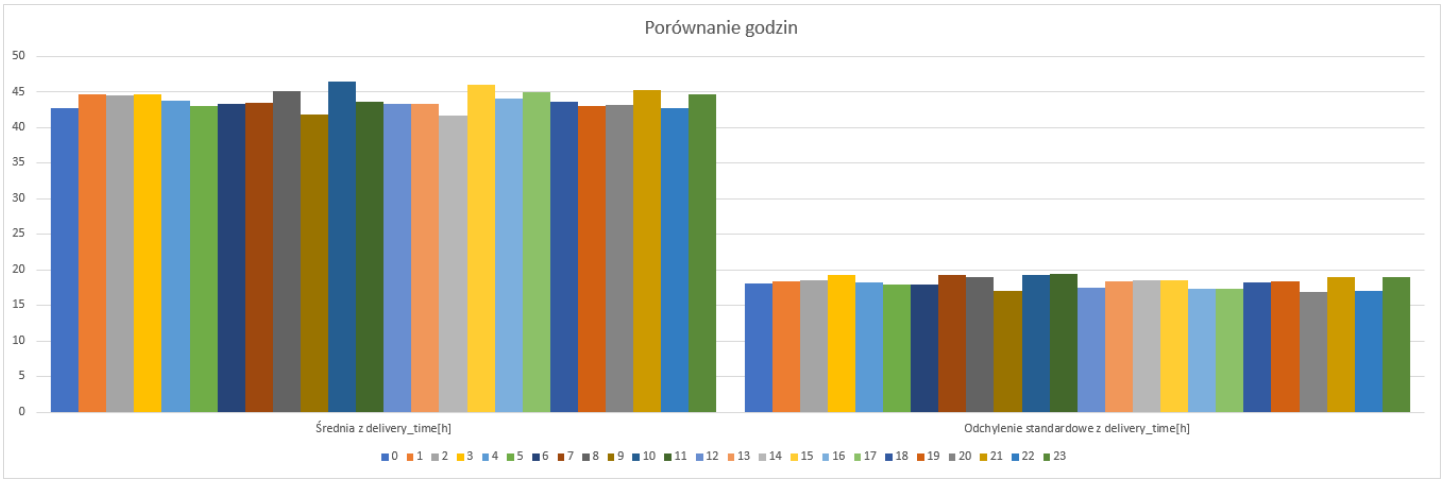
Dzień miesiąca:



Tutaj lekka rozbieżność średnich wynika z wpływu dnia tygodnia na czas transportu.

Specjalne dni, np. święta:
W święta przesyłki też są dostarczane i średnia czasów dostaw dla świąt jest bardzo zbliżona do ogólnej.
Nie sprawdzaliśmy rozkładu dla każdego dnia osobno, ale nie natknęliśmy się na żadne odstające dni lub większe skupiska dłuższych dostaw w pewnych okresach czasu.

Godzina zakupu:



Założenia:

- Zakupione produkty nie ulegają “zagubieniu” w trakcie transportu, zawsze docierają do klienta.
- Firmy kurierskie nie pracują w niedziele.
- Dostawy odbywają się święta tak samo jak w normalne dni.
- Dostawy odbywają się jedynie w godzinach 8.00-19.00, niezależnie od firmy transportowej.
- Model będzie przewidywał czasy transportu dla ograniczonej liczby miast i firm transportowych
- Dane zawsze będą kompletne, tak jak te, które otrzymaliśmy (niekompletne rekordy mogą być odrzucane)

Kompletność i reprezentatywność danych:

W danych nie stwierdzono brakujących atrybutów ani wyraźnych sprzeczności (wszystkie rekordy zawierają wszystkie pola, czasy dostaw nie są ujemne).

Do pełnego obrazu brakuje jednak dostaw dokonanych od końca października do końca roku. Mimo, że nie uwzględniamy ich w modelu, mogłoby się okazać, że powinniśmy, ponieważ akurat w tym czasie coś się zmienia (np. pojawiają się opóźnienia związane z liczbą zamówień przed świętami).

Żeby model dostosowywał się lepiej do zjawisk zmieniających się na przestrzeni lat, przydatna byłaby również chociaż niewielka ilość, najlepiej równo rozłożonych, danych z kilku różnych lat ubiegłych.

Ostatecznie jednak dane, dla podanej sytuacji, uznajemy za wystarczające do skonstruowania modelu spełniającego wyznaczone przez nas kryteria. Model może jednak nie być w stanie dostosować się do zmian w ciągu kolejnych lat i będzie ograniczony do tych miast i firm kurierskich, które występują obecnie w danych.