

Raport: Analiza Predykcji Czasu Dostawy

1. Wprowadzenie

Nasza firma przewiduje czasy dostawy na podstawie prostego modelu, który wykorzystuje średnią ze wszystkich zarejestrowanych danych. Ostatnie spostrzeżenia kierowców oraz nasze analizy wskazały, że prognozowany czas dostawy znacząco odbiega od rzeczywistego. Dodatkowo wiadomo, że dostawy do domów jednorodzinnych odbywają się znacznie szybciej niż do budynków wielorodzinnych. Niestety, ponieważ nie posiadamy bezpośrednich danych o typie budynku, postanowiliśmy zbadać, czy inne dostępne zmienne (np. sektor, godzina, dzień tygodnia, liczba produktów) mogą służyć jako wskaźniki różnych warunków operacyjnych.

W niniejszym raporcie przedstawiamy nasze główne założenia, przeprowadzoną metodologię oraz wyniki analizy, ilustrowane odpowiednimi wykresami.

2. Metodologia i Przyjęte Założenia

Założenia:

- **Planowany czas dostawy:** Aktualnie ustalany jest na stałą wartość około 175–180 sekund (czyli około 3 minut), niezależnie od specyfiki zamówienia.
- **Brak danych o typie budynku:** Nie posiadamy bezpośrednio informacji, czy zamówienie realizowane jest do domu jednorodzinnego czy do budynku wielorodzinnego. Dlatego przyjmujemy, że inne cechy (np. sektor, liczba produktów, waga zamówienia) mogą być pomocne w różnicowaniu procesów.
- **Czystość danych:** Przed analizą usunięto rekordy z brakującymi danymi oraz te, które miały nieprawidłowe (np. ujemne) wartości czasu dostawy.

Proces przygotowania i analizy danych:

1. **Wczytanie i czyszczenie danych:** Dane z GPS (segmenty trasy), zamówień, produktów oraz ich połączeń zostały wczytane, a rekordy z brakującymi lub błędnymi danymi usunięte.
2. **Ekstrakcja segmentów typu „STOP”:** Skupiliśmy się na segmentach, w których nastąpiło zatrzymanie, obliczając rzeczywisty czas dostawy jako różnicę między czasem rozpoczęcia a zakończenia segmentu. Wynik został przeliczony na sekundy oraz zaokrąglony do minut (z dokładnością 1-minutową).
3. **Obliczenie błędu predykcji:** Błąd predykcji definiujemy jako różnicę między rzeczywistym czasem dostawy a planowanym czasem. Zaobserwowaliśmy, że planowany czas (około 3 minut) jest bardzo niski w porównaniu do rzeczywistych czasów (około 240–250 minut).
4. **Wizualizacja danych i agregacje:** Sporządziliśmy histogram rzeczywistego czasu dostawy z dokładnością do jednej minuty, Sporządziliśmy histogram błędu predykcji (różnica czasu rzeczywistego i planowanego), Przeanalizowaliśmy średnie czasy

dostaw wg sektorów, godzin oraz dni tygodnia, Grupowaliśmy dane (w tym zamówienia z błędem powyżej 100 minut) oraz wygenerowaliśmy tabele zbiorcze i wykresy, które mają pomóc wyłapać ewentualne korelacje.

3. Wyniki Analizy

3.1. Histogram rzeczywistego czasu dostawy

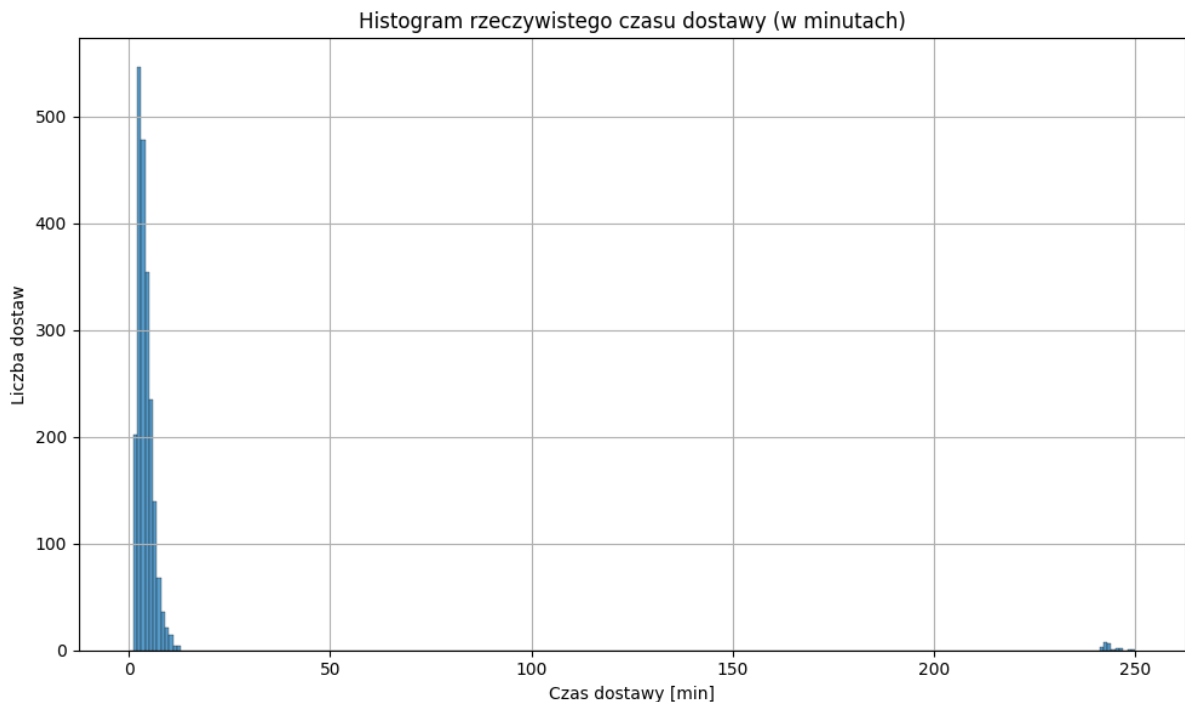


Chart 1: Histogram rzeczywistego czasu dostawy (w minutach)

Odpowiedź na pytanie 1: Na pierwszym wykresie („Histogram rzeczywistego czasu dostawy (w minutach)”) widzimy rozkład faktycznych czasów dostaw.

- **Co pokazuje:** Niektóre dostawy trwają od około 240 do 250 minut, co jest znacznie wyższe od planowanego czasu (około 3 minuty).
- **Wniosek:** Użycie jednej średniej wartości do prognozowania jest niedokładne, bowiem nie uwzględnia różnych warunków operacyjnych.

3.2. Histogram błędu predykcji

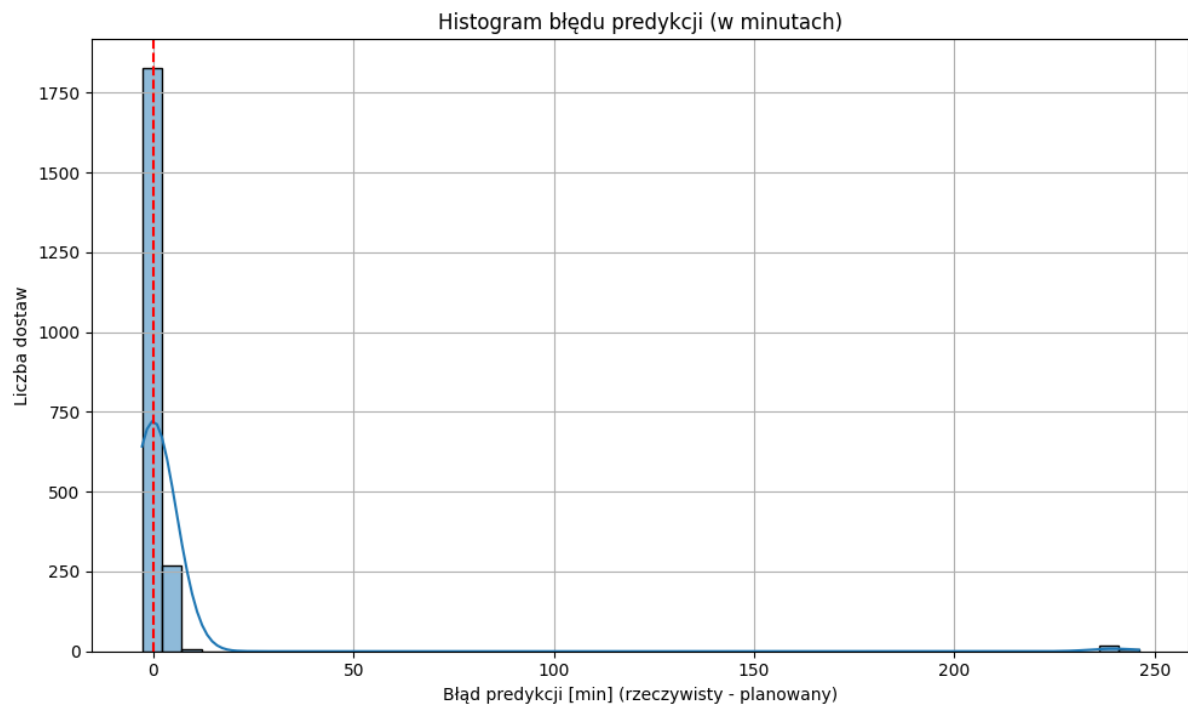


Chart 2: Histogram błędu predykcji (w minutach)

Odpowiedź na pytanie 2: Drugi wykres („Histogram błędu predykcji (w minutach)”) przedstawia rozkład różnicy między rzeczywistym a planowanym czasem dostawy.

- **Co pokazuje:** Wartość błędu predykcji wynosi średnio około 240 minut dla niektórych przypadków, co oznacza, że planowany czas jest drastycznie заниżony.
- **Wniosek:** Aktualny sposób prognozowania (oparty jedynie na średniej) nie daje wiarygodnej prognozy, stąd model musi zostać udoskonalony.

3.3. Średni czas dostawy wg sektorów

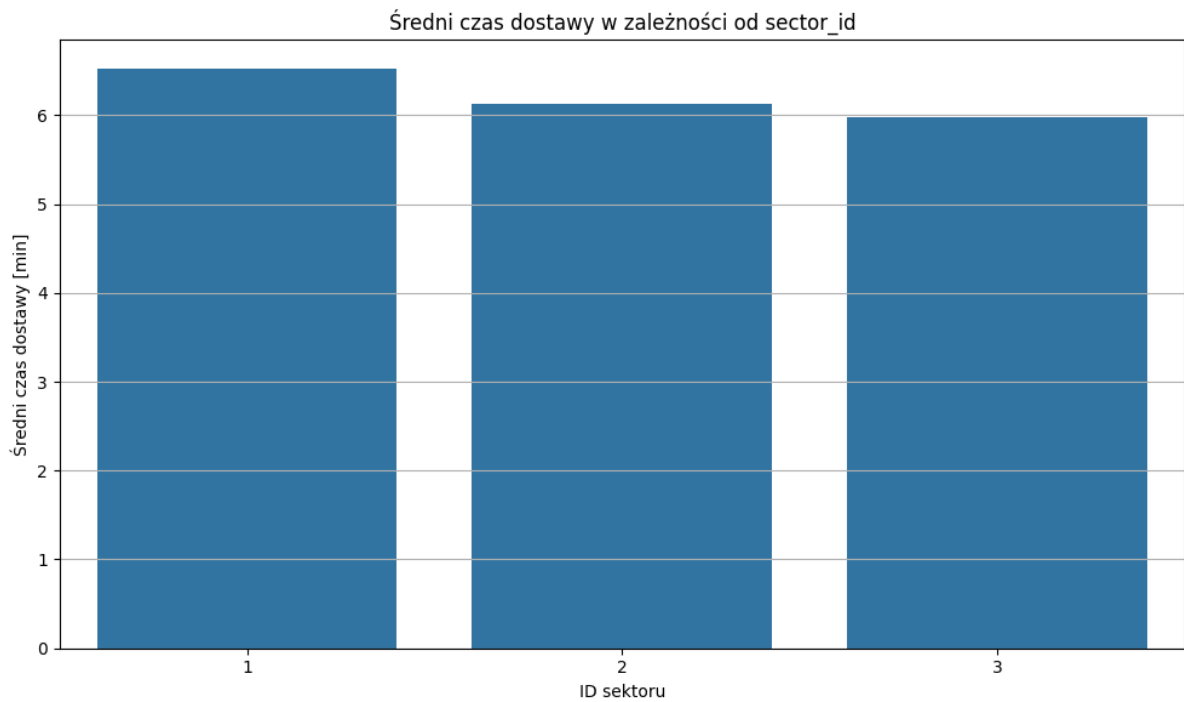


Chart 3: Średni czas dostawy wg sector_id

Odpowiedź na pytanie 3: Kolejny wykres („Średni czas dostawy w zależności od sector_id”) pokazuje, że średni czas dostawy różni się między sektorami.

- **Co pokazuje:** Choć różnice nie są ogromne, zauważalne są systematyczne odchylenia, niektóre sektory wykazują nieco wyższy czas dostawy.
- **Wniosek:** Może to świadczyć, że zamówienia realizowane w określonych regionach (które mogłyby odpowiadać dostawom do budynków wielorodzinnych) są trudniejsze i wolniejsze do obsłużenia. Włączenie cechy „sector_id” do modelu predykcyjnego mogłoby pozwolić na lepsze dostosowanie prognozy.

3.4. Agregacje i dodatkowe analizy

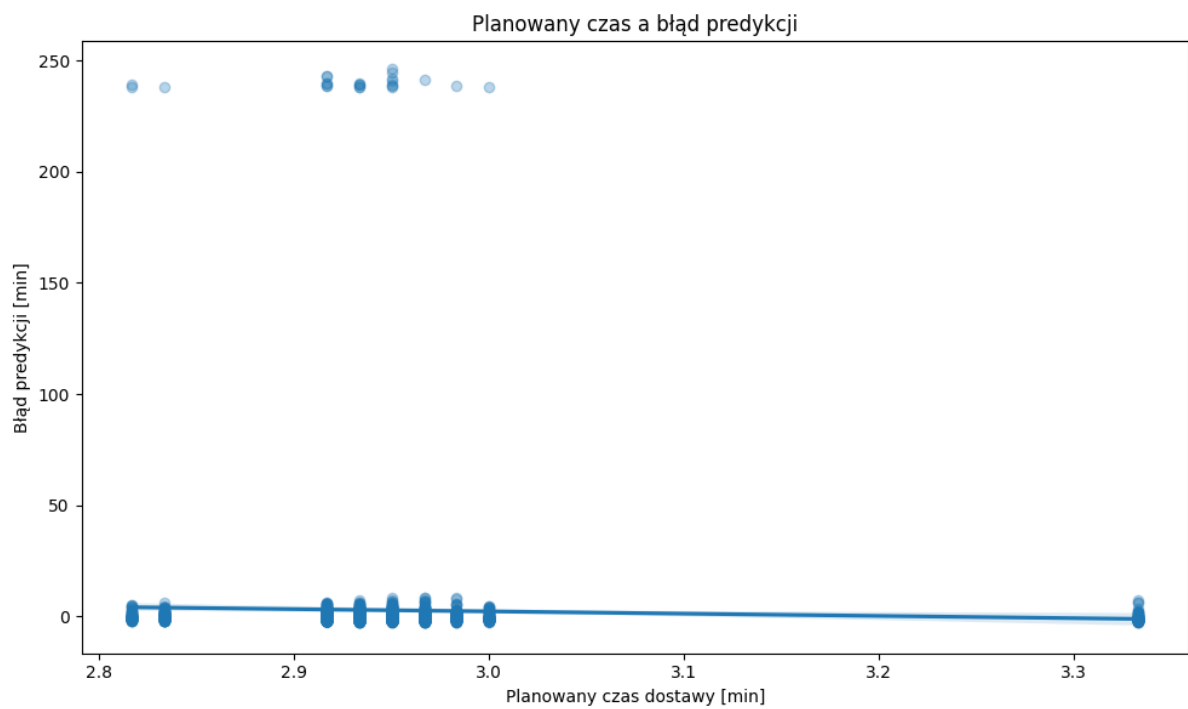
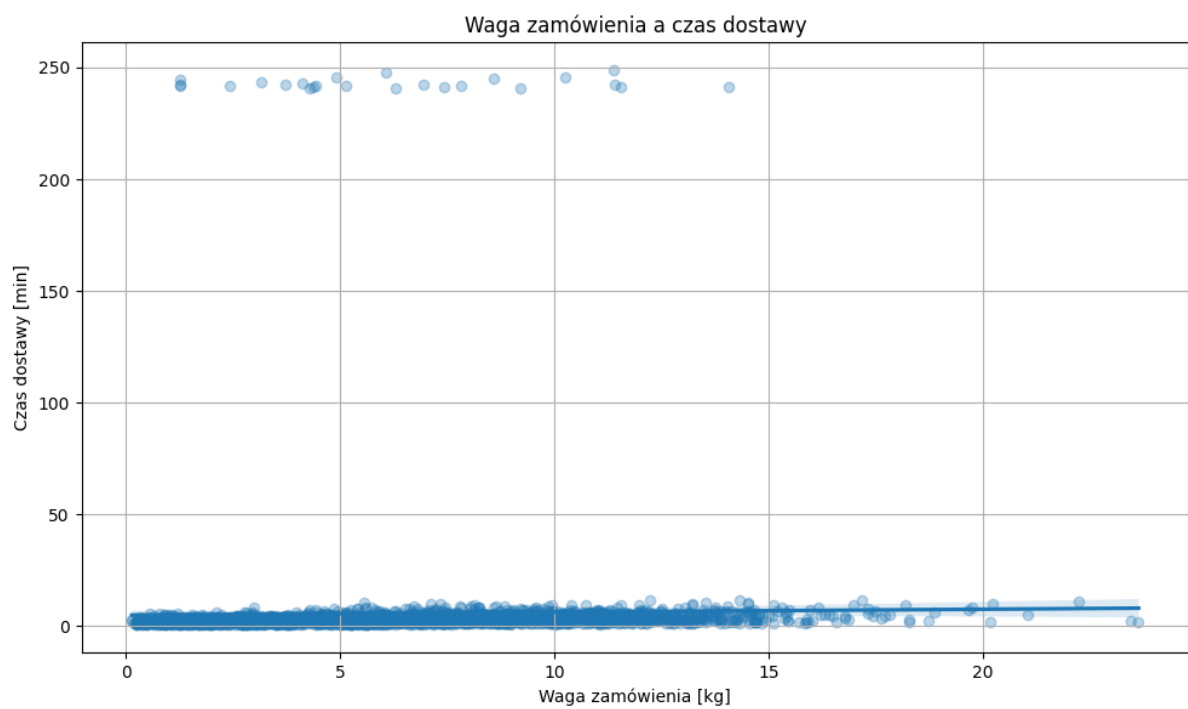
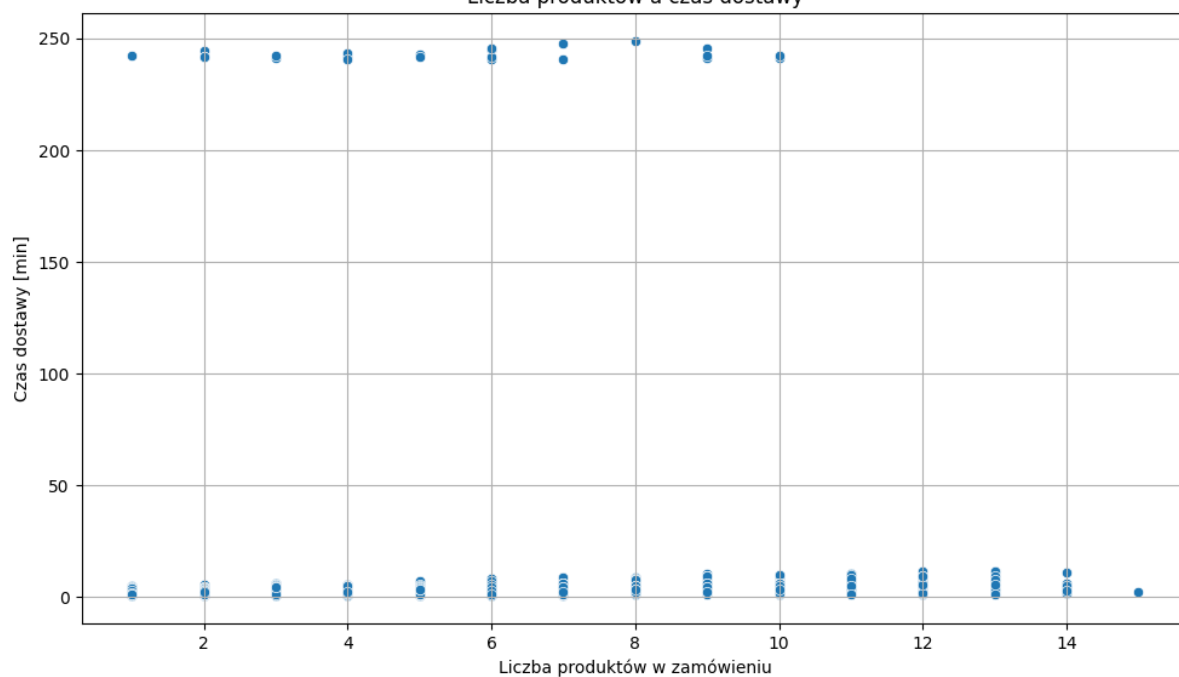


Chart 4: Planowany czas a błąd predykcji



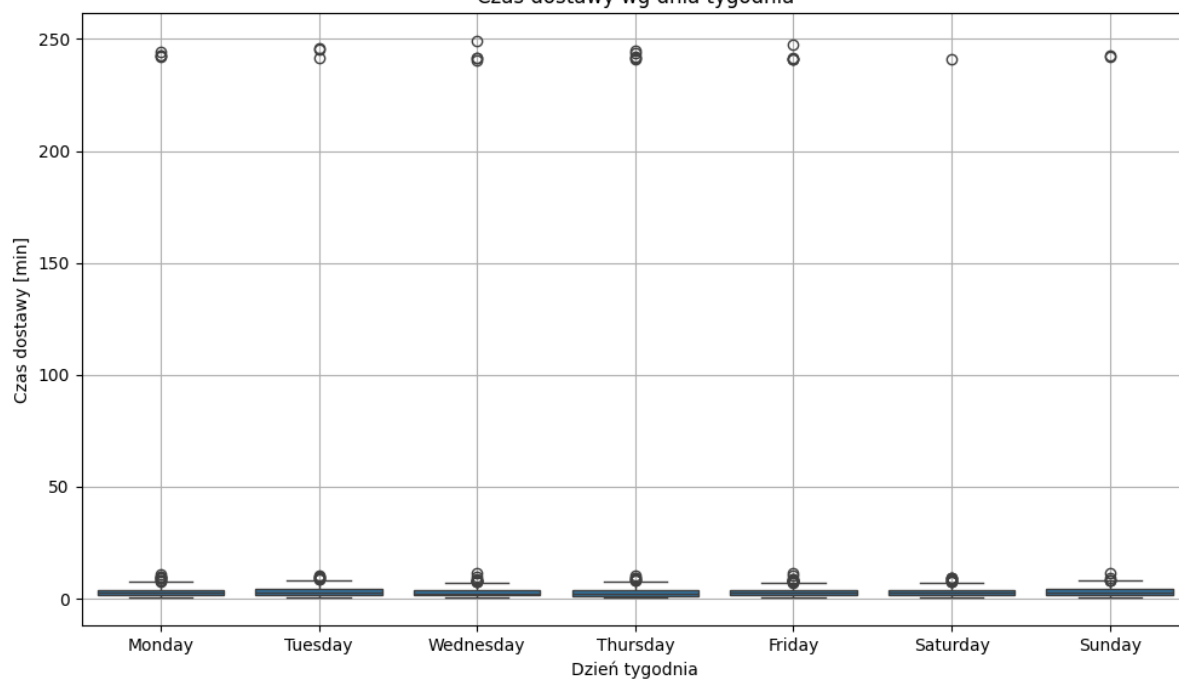
Dodatkowe wykresy

Liczba produktów a czas dostawy

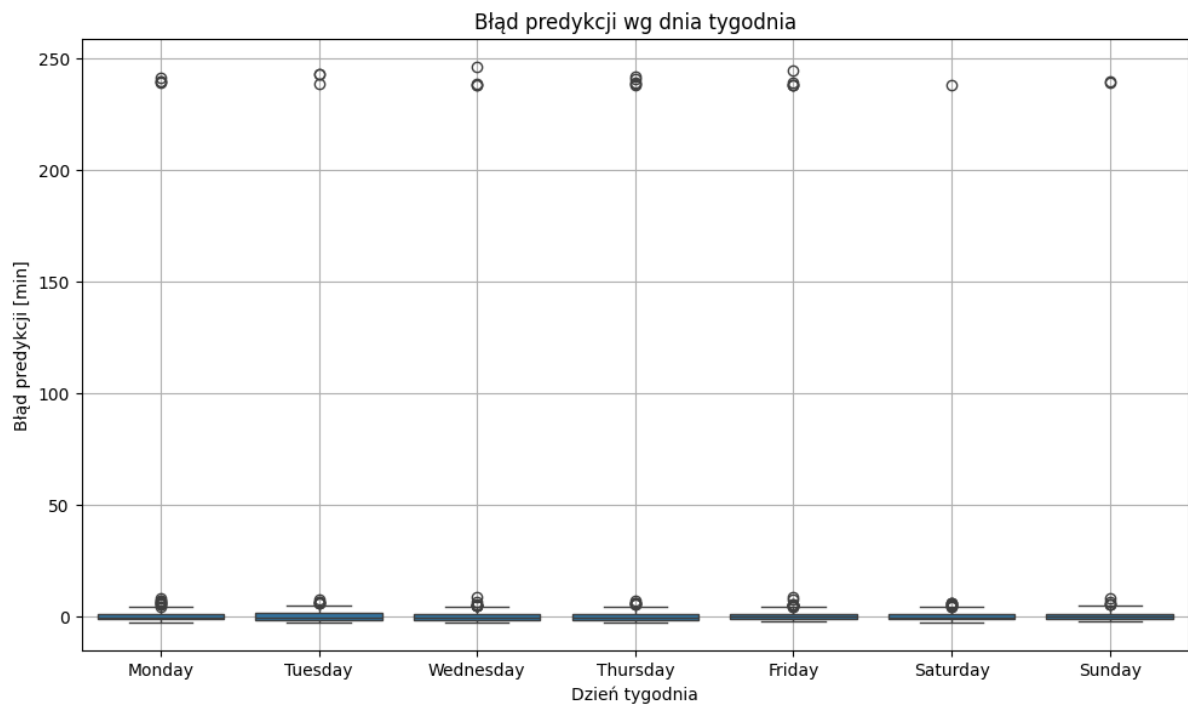


Dodatkowe wykresy

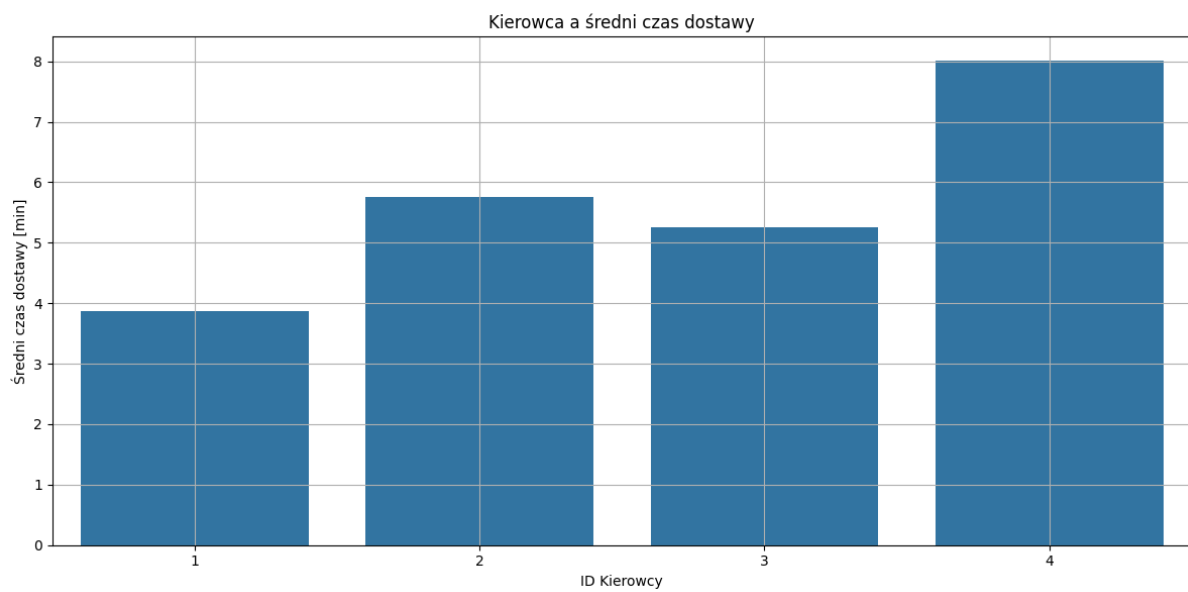
Czas dostawy wg dnia tygodnia



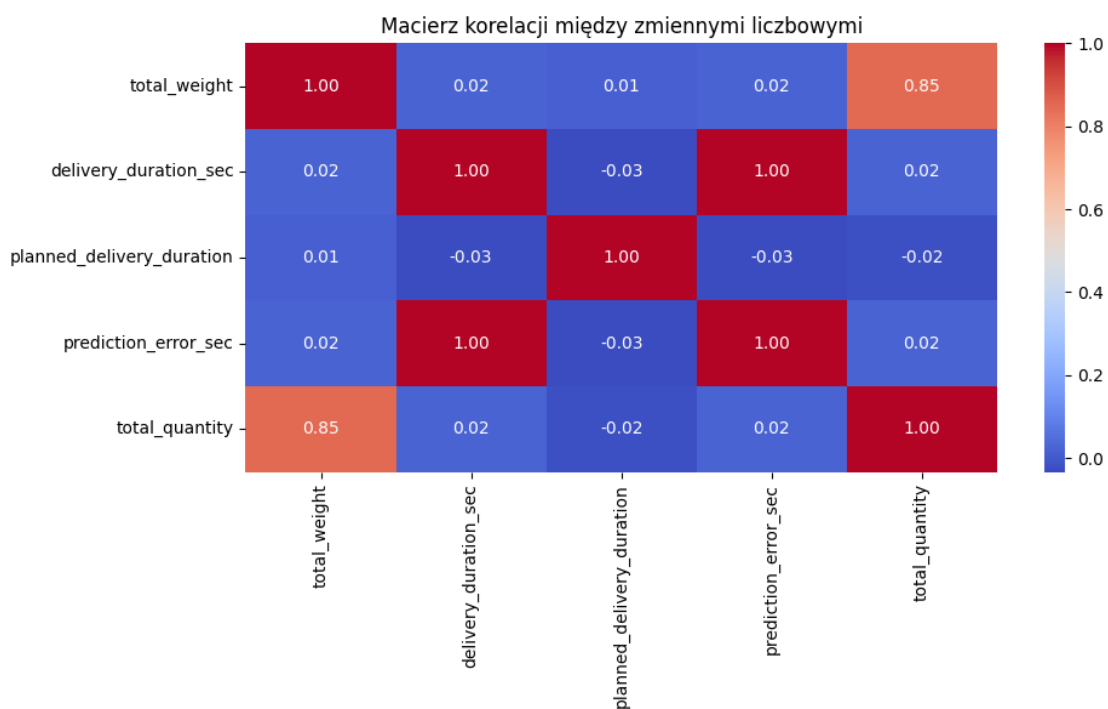
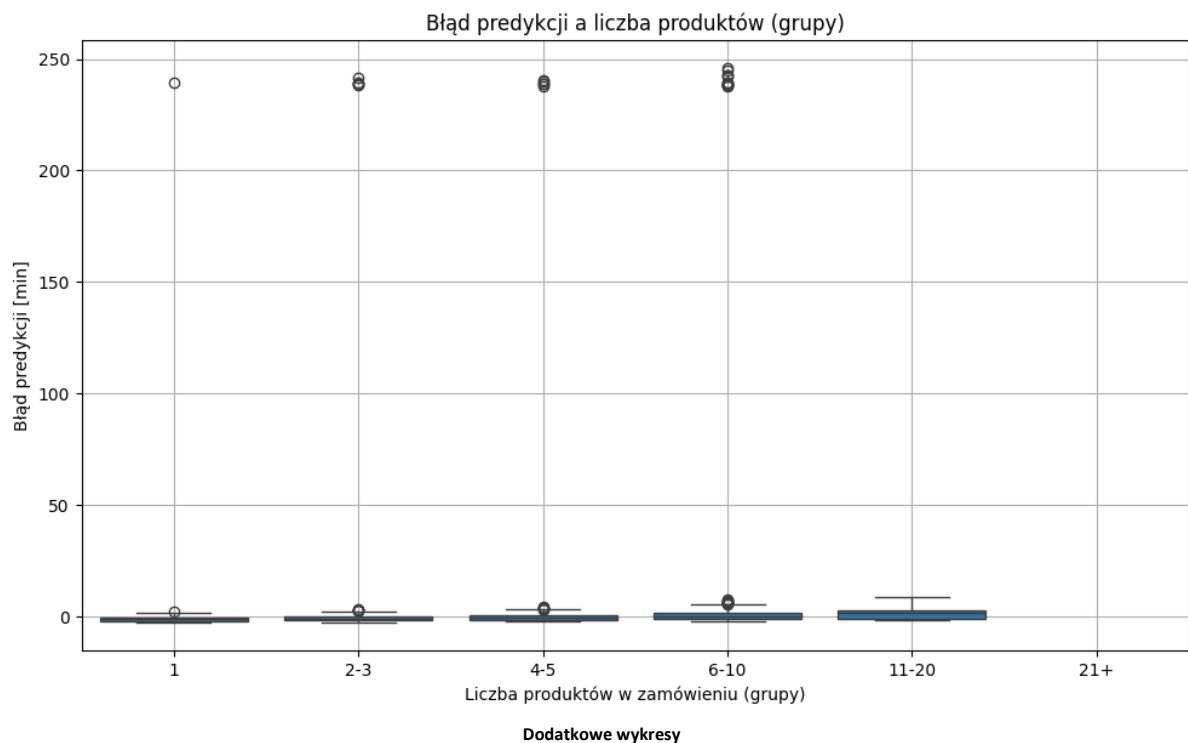
Dodatkowe wykresy



Dodatkowe wykresy



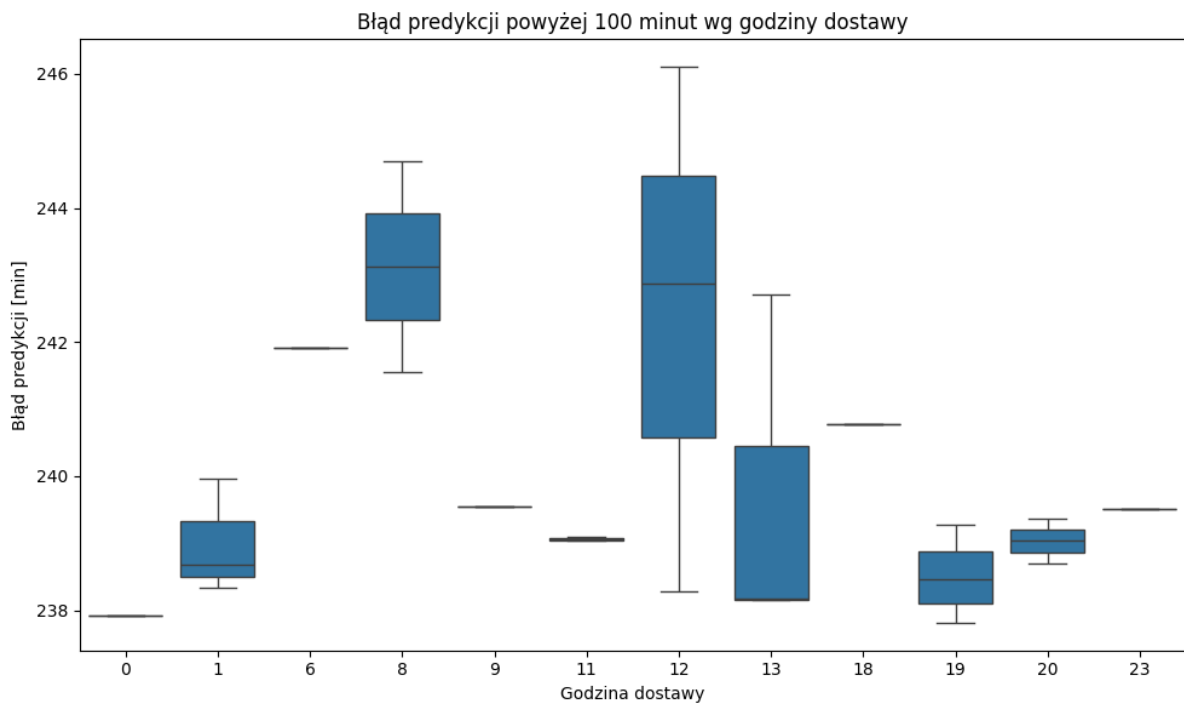
Dodatkowe wykresy

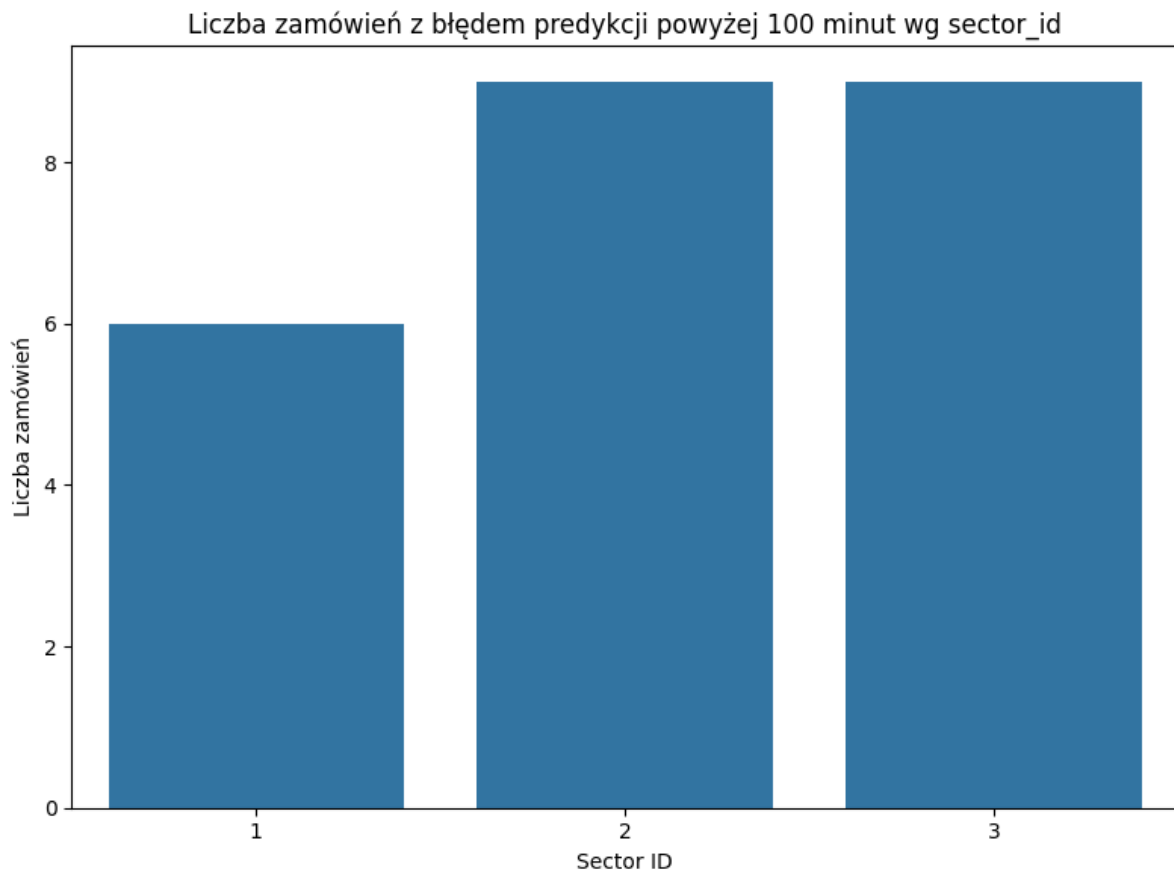


Odpowiedź na pytanie 4: Na podstawie wielu wykresów (np. „Planowany czas a błąd predykcji”, „Waga zamówienia a czas dostawy”, „Liczba produktów a czas dostawy”, „Godzina zakończenia dostawy a błąd predykcji”, „Czas dostawy wg dnia tygodnia” oraz „Błąd predykcji wg dnia tygodnia”) dokonaliśmy gruntownej analizy, która wykazała:

- **Brak elastyczności modelu:** Planowany czas dostawy jest ustalany na wartość około 3 minut, niezależnie od wielkości, wagi, liczby produktów, a także pory dnia czy dnia tygodnia.
- **Niedoszacowanie czasu:** Rzeczywisty czas dostawy niektórych przypadków waha się między 240 a 250 minutami, co skutkuje bardzo wysokim błędem predykcji (średnio ponad 238 minut).
- **Potencjalne korelacje:** Analiza korelacji wykazała, że cechy takie jak waga zamówienia i liczba produktów są ze sobą silnie powiązane (współczynnik około 0.85), ale nie wyjaśniają w pełni rozbieżności między planowanym a rzeczywistym czasem. Agregacja danych według sektora, godziny oraz dnia tygodnia nie ujawniła drastycznych różnic, ale systematyczne, choć drobne odchylenia są widoczne.

Na wykresach ukazanych poniżej analizowaliśmy zamówienia, dla których błąd predykcji przekracza 100 minut.





Pomimo tego, że ilość zamówień o wysokim współczynniku błędu wynosi 24, to, ich analiza potwierdza, że:

- Wszystkie charakteryzują się niemal identycznymi bardzo niskimi czasami planowanymi (około 175–180 sekund),
- Rzeczywisty czas dostawy jest znacznie wyższy (około 14400–14900 sekund),
- Średni błąd wynosi ponad 238 minut niezależnie od pory dnia, sektora czy dnia tygodnia.

Wniosek: Wyniki naszej analizy wskazują, że metoda oparta na jednej średniej wartości nie uwzględnia złożoności procesu dostawy. W związku z tym rekomendujemy:

- **Segmentację zamówień:** Podział na grupy (np. na podstawie sektora, liczby produktów, wagi zamówienia) pozwoli na wprowadzenie korekt do prognozy.
- **Wdrożenie bardziej zaawansowanego modelu:** Rozważenie algorytmów regresyjnych, drzew decyzyjnych lub modeli uczenia maszynowego, które będą mogły lepiej uchwycić zależności między zmiennymi, a tym samym poprawić jakość prognozy.

4. Podsumowanie

Nasza analiza wykazała, że:

- **Aktualny model:** Oparty wyłącznie na średniej powoduje drastycznie niedoszacowany planowany czas dostawy (około 3 minut w porównaniu do rzeczywistych 240–250 minut).
- **Różnice między sektorami:** Choć niewielkie, wskazują, że nie wszystkie obszary operują w ten sam sposób, co może być proxy dla różnic między dostawami do domów jednorodzinnych a dostawami do budynków wielorodzinnych.
- **Dodatkowe zmienne:** Analiza wagi, liczby produktów, godziny oraz dnia tygodnia sugeruje, że istnieją subtelne zależności, które mogłyby zostać wykorzystane do lepszego skalibrowania modelu.

Rekomendujemy zmianę podejścia i wdrożenie modelu predykcyjnego, który uwzględni dodatkowe zmienne i segmentację zamówień, dzięki czemu prognozy staną się bardziej precyzyjne, a błędy istotnie mniejsze.

5. Załączniki (Wykresy)

- **Chart 1: Histogram rzeczywistego czasu dostawy (w minutach):** Pokazuje, że niektóre dostawy trwają od 240 do 250 minut.
- **Chart 2: Histogram błędu predykcji (w minutach):** Ilustruje, że różnica między planowanym a rzeczywistym czasem wynosi w niektórych przypadkach średnio około 240 minut, co jest bardzo wysokim błędem.
- **Chart 3: Średni czas dostawy wg sector_id:** Wykres słupkowy ukazuje różnice w średnim czasie dostawy między sektorami, co może być wykorzystane jako proxy dla różnic w typie budynków.
- **Chart 4: Planowany czas a błąd predykcji:** Wykres regresji pokazuje, że niska wartość planowanego czasu powoduje bardzo wysoki błąd predykcji.
- **Dodatkowe wykresy:** Wykresy pokazujące zależność czasu dostawy od wagi zamówienia oraz liczby produktów: Boxploty ilustrujące wpływ godziny dostawy i dnia tygodnia na błąd predykcji, Macierz korelacji między zmiennymi liczbowymi, która pomaga uchwycić zależności między poszczególnymi cechami zamówień.
- **Wykresy danych o wysokim współczynniku błędu:** Wykresy ukazujące wartości rekordów o wysokim współczynniku danych, pokazujące zależności od np. dni tygodnia, czy danego sektora.

Raport ten podsumowuje nasze dotychczasowe ustalenia i pokazuje, że jednolita metoda prognozowania (oparta na średniej) nie oddaje rzeczywistości. Przyjęcie bardziej złożonego modelu oraz segmentacja danych mogą znacząco poprawić jakość prognozy czasu dostawy.