

Problem gapowicza w transporcie publicznym

Autorzy: Sebastian Halicki, Zofia Koziel, Lizaveta Miastouskaya

2023

Spis treści

1	Podsumowanie	3
2	Opis organizacji	3
3	Opis problemu	3
4	Wyniki analizy	4
4.1	Dane wejściowe	4
4.2	Rozwiązanie optymalne	5
5	Analiza Wrażliwości	7
5.1	Analiza wrażliwości ze względu na cenę biletu	7
5.2	Analiza wrażliwości ze względu na koszt kontroli	8
6	Model regresji	9
6.1	Badanie wpływu zmiennych na średnie przychody	9
6.2	Model regresji liniowej	11
6.3	Wyniki modelu	11
7	Wnioski i zalecenia	13
8	Bibliografia	13

1 Podsumowanie

Poniższy tekst omawia problem gapowicza i jego różne aspekty. Zwraca się uwagę na średnie przychody oraz ich zmienność, która różni się w zależności od wartości zadanych parametrów - ceny biletu, ceny mandatu, liczby przeprowadzonych kontroli w autobusach oraz kosztu owych kontroli. W pierwszej kolejności przedstawione jest rozwiązanie optymalne dla zadanych parametrów modelu. Przeprowadzona symulacja pozwoliła na określenie rozwiązań pareto optymalnych - rozwiązaniach z jednocześnie największymi średnimi przychodów i najmniejszymi odchileniami standardowymi przychodów stanowiących miarę ryzyka. Jako rozwiązanie optymalne ustalono wysokość kary za brak biletu na 80 PLN przy liczbie kontroli biletowych w ciągu dnia równej 600. Następnie przedstawione są zmiany rozwiązania optymalnego w zależności od ceny biletu oraz kosztu kontroli. Ostatnim krokiem jest konstrukcja modelu regresji liniowej, która pokazuje jak i w jakim stopniu cena biletu wpływa na średni przychód.

2 Opis organizacji

Niniejszy raport został wykonany na zlecenie Zarządu Komunikacji Miejskiej w związku ze zmniejszeniem odsetka osób kupujących bilety wśród użytkowników komunikacji miejskiej. W odpowiedzi na rosnącą liczbę pasażerów korzystających z usług transportu zbiorowego, istnieje potrzeba efektywnego zarządzania tym problemem w celu zapewnienia stabilności finansowej. Optymalizując liczbę kontrolerów i wysokość kary za brak biletu, organizacja będzie mogła skuteczniej radzić sobie z problemem niepłacących pasażerów, minimalizując związane z tym straty i maksymalizując zysk. Podobny problem jest omawiany w pracy Basiglio S. pod tytułem "Free riders in Public Transport" z 2013 roku.

3 Opis problemu

W ostatnich miesiącach komunikacja miejska odnotowała zmniejszenie przychodu ze sprzedaży biletów przy zachowaniu porównywalnej liczby pasażerów wskazuje na wzrost liczby niepłacących pasażerów. Jako możliwe przyczyny takiej sytuacji wskazano niedostateczną liczbę kontroli i nieodpowiednią wartość kary pieniężnej za przejazd bez ważnego biletu.

Analiza przedstawionego problemu ma na celu znalezienie optymalnej liczby kontroli i wysokości kary za brak biletu, które pozwolą zmaksymalizować zysk z przejazdów autobusowych. Problem gapowicza pojawia się w literaturze naukowej istnieją gotowe modele (Basiglio.S. "Free Riders in Public Transport"), na potrzeby analizy problemu zdecydowano się na skonstruowanie nowego modelu zysków z kursów autobusowych. Każdego dnia zakłada się tę samą liczbę przejazdów, dla każdego z nich, z rozkładu Poissona losowana jest liczba pasażerów x_i ograniczona przez pojemność autobusu (jednostkowy koszt pasażera oznaczono jako k_p). Dla każdego pasażera zachodzi losowanie czy zdecyduje się na jazdę bez biletu. Przyjęto założenie o racjonalności

pasażerów dlatego prawdopodobieństwo przejazdu bez biletu ustalono jako:

$$p_{jazdybezbiletu} = \frac{cena\ biletu}{cena\ biletu + kara}$$

przy takiej konstrukcji w przypadku braku kary pasażerowie przestają kupować bilety. W dalszej kolejności w przypadku dostępnych kontrolerów dochodzi do kontroli danego kursu o ustalonym koszcie płaconym przez Zarząd Komunikacji. W trakcie kontroli dla dostatecznie małej liczby pasażerów wszyscy bez biletów zostają złapani, jeśli pasażerów jest więcej, gapowicze mają możliwość schować się wśród innych pasażerów przed kontrolą. Wszyscy złapani gapowicze płacą taką samą karę. Funkcja celu wyznaczająca zysk ma postać:

$$F = \sum_{j=1}^{n_{dni}} \left(\sum_{i=1}^{n_{kursu}} z_{i,j} kara + (x_{i,j} - fr_{i,j})(cena\ biletu) - x_{i,j} k_p - k_{kontroli} - k_{kursu} \right)$$

gdzie:

- $z_{i,j}$ - liczba złapanych pasażerów bez biletu w i -tym kursie, j -ego dnia
- $kara$ - wartość kary opłacanej przez złapanego gapowicza
- $fr_{i,j}$ - liczba pasażerów bez biletu w i -tym kursie, j -ego dnia
- k_p - koszt jednostkowy dodatkowego pasażera
- k_{kursu} - koszt wykonania jednego kursu
- $k_{kontroli}$ - koszt kontroli biletowej jeśli została przeprowadzona.

Szukanym rozwiązaniem będzie zatem

$$(\bar{kara}, \bar{n}_{kontroli}) = \arg \max_{kara, n_{kontroli} > 0} F$$

4 Wyniki analizy

4.1 Dane wejściowe

Na potrzeby analizy przyjęto następujące założenia odnośnie wartości bazowych:

- Cena biletu: 5 PLN
- Maksymalna liczba pasażerów kursu: 60
- Średnia liczba osób chętnych na dany kurs: 20
- Liczba kursów w ciągu dnia: 100
- Koszt przejazdu bez pasażerów: 100 PLN
- Dodatkowy koszt związany z przewozem pasażera: 3.5 PLN

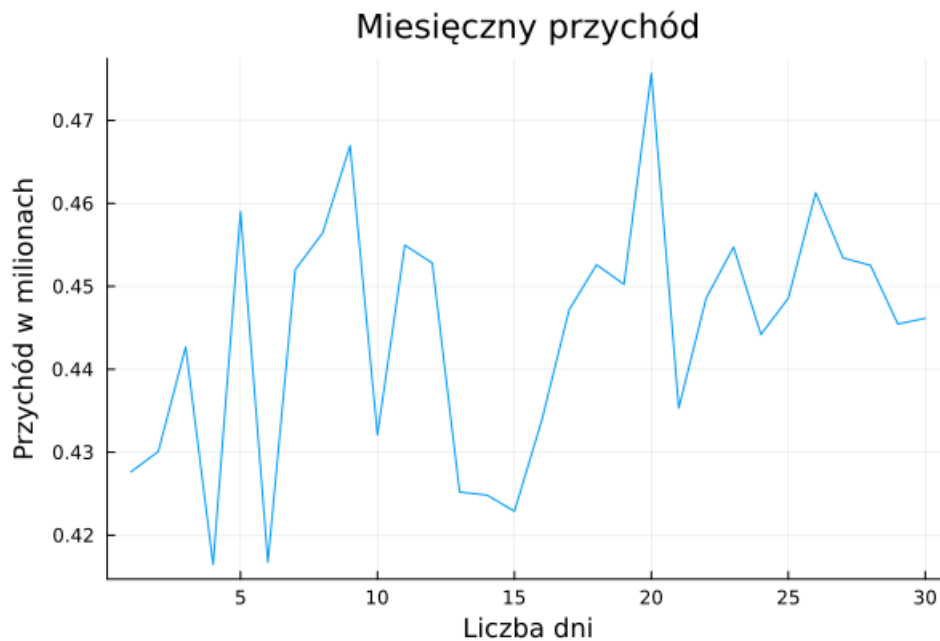
- Koszt przeprowadzenia kontroli kursu: 20 PLN
- Liczba pasażerów, których kontrola na pewno sprawdzi: 30

Prawdopodobieństwo, że pasażer będzie gapowiczem została ustalona jako stosunek ceny biletu do sumy ceny biletu i mandatu. Wykorzystanie rozkładu Poissona, pozwala symulacji uwzględniać losowość liczby pasażerów w różnych przejazdach autobusowych, dając bardziej realistyczną symulację systemu komunikacji miejskiej.

4.2 Rozwiązanie optymalne

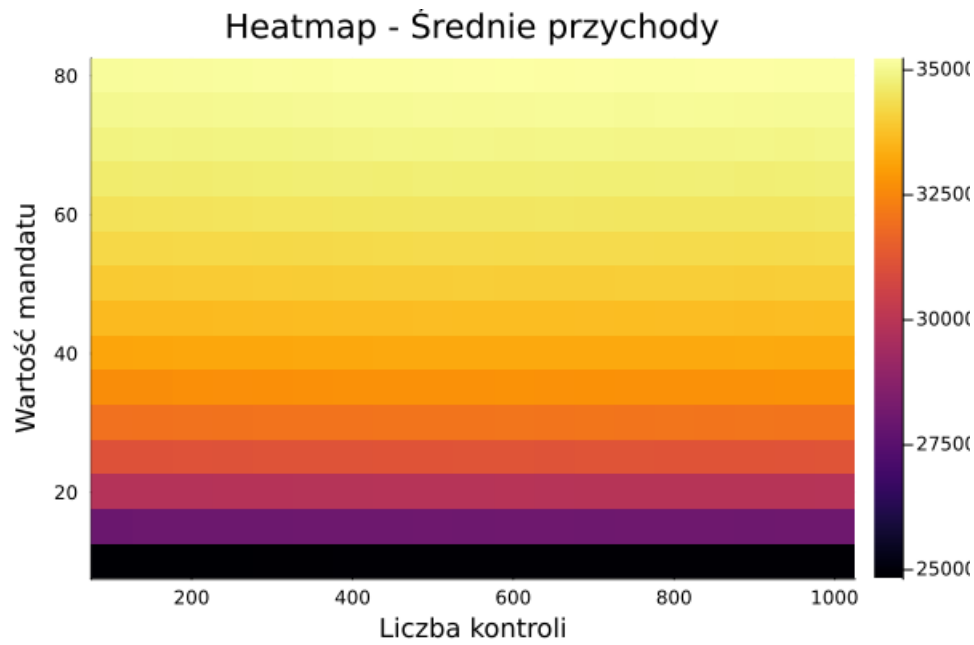
W wyniku analizy zidentyfikowano funkcję, która symuluje jeden dzień działalności transportowej oraz funkcję, która prowadzi symulację przez 30 dni, która sumuje przychód z każdego dnia i zwraca łączny przychód oraz wektor przychodu dla poszczególnych dni.

Z wykresu poniżej można zauważyć, że w ciągu miesiąca dochodzi do pewnych wahań dziennego zysku jednak jego wartości zawierają się w przedziale $(0.4, 0.48)$. Można założyć, że różnice wynikają z czynników takich jak zmienność liczby pasażerów, kontrola biletów, koszty operacyjne i inne czynniki przychodowe.



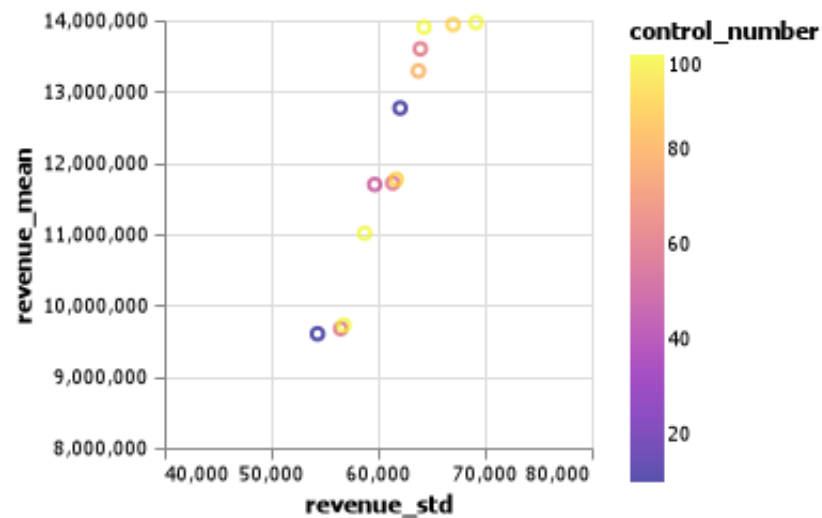
Rysunek 1: Dzienny przychód działalności transportowej w ciągu miesiąca

Przestrzeń na której dokonywano optymalizacji ustalono jako siatkę gdzie $kara$ należała do przedziału $[10, 80]$ z przejściami co 5, a $n_{kontroli}$ do przedziału $[10, 100]$ z przejściami co 10. Dla każdego układu wykonano 500 symulacji i wyciągnięto średni zysk oraz odchylenie standardowe. Poniższa wizualizacja za pomocą heatmapy sugeruje, że rozwiązania optymalne będzie zachodzić dla najwyższych wartości kary, jednak trudno ocenić na tej podstawie liczbę kontroli, która ma znacznie mniejszy wpływ na optymalne rozwiązanie.

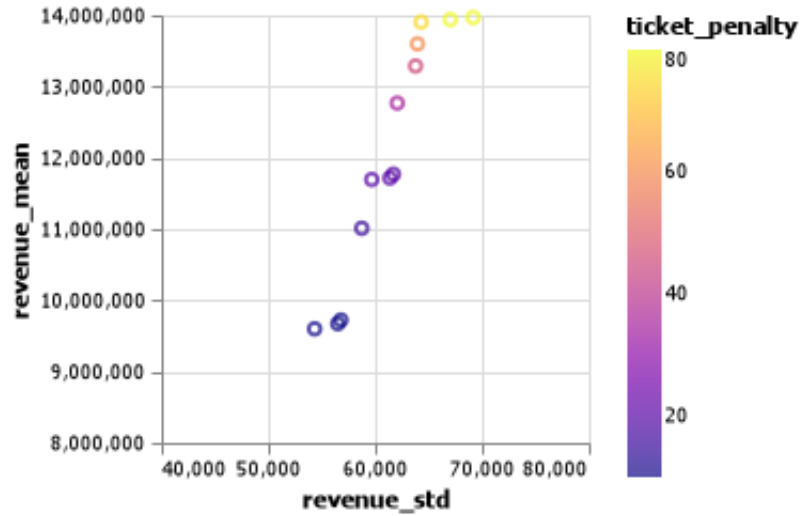


Rysunek 2: Średni zysk w zależności od mandatu i liczby kontroli

Przy przyjętych założeniach optymalnym rozwiązaniem jest $(kara, n_{kontroli}) = (80, 100)$, co może wynikać ze stosunkowo niskiego kosztu kontroli względem zysku ze złapania pasażera na gapę przez co nawet duża liczba kontroli nie jest tak wyraźnie odczuwalna dla zysków.



Rysunek 3: Rozwiązania pareto optymalne - liczba kontroli



Rysunek 4: Rozwiązania pareto optymalne - koszt mandatu

5 Analiza Wrażliwości

W ramach analizy wrażliwości został zbadany wpływ ceny biletu oraz kosztu kontroli na średni przychód miesięczny komunikacji miejskiej.

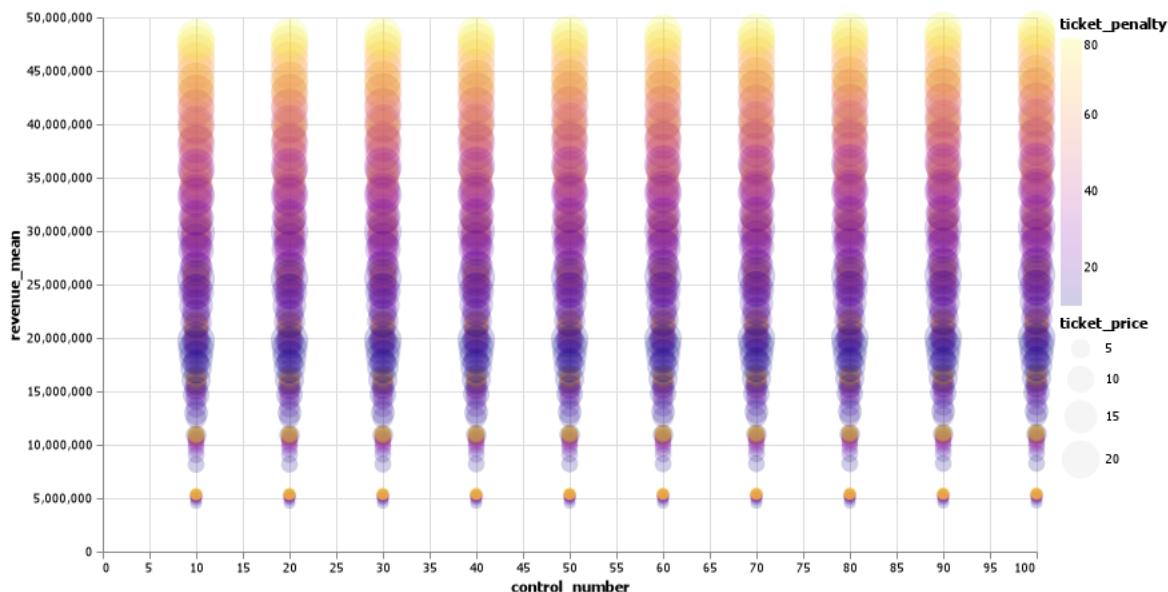
5.1 Analiza wrażliwości ze względu na cenę biletu

W tabeli 1 przedstawione zostały wyniki optymalne (maksymalne przychody) dla poszczególnych cen biletów, oraz odpowiadające im wartości liczby kontroli. Najwyższe średnie przychody są osiągnięte dla ceny biletu 20PLN oraz liczby kontroli 100.

Liczba kontroli	Cena biletu	Koszt mandatu	Średnie przychody
100	2.0	80.0	5.45859e6
100	4.0	80.0	1.11966e7
100	6.0	80.0	1.66637e7
100	8.0	80.0	2.18928e7
100	10.0	80.0	2.68872e7
100	12.0	80.0	3.16618e7
100	14.0	80.0	3.62254e7
100	16.0	80.0	4.06248e7
100	18.0	80.0	4.48e7
100	20.0	80.0	4.88483e7

Tabela 1: Rozwiązania optymalne dla poszczególnych cen biletów

Na wykresie 5 możemy zauważyć jak kształtują się zależności czterech zmiennych. Przede wszystkim widać na nim, że cena wyższy koszt mandatu warunkuje wyższy średni przychód. Możemy zaobserwować również, że średni przychód nie zależy od liczby kontroli, natomiast mniejsza cena biletu implikuje mniejszy średni przychód.



Rysunek 5: Wykres bąbelkowy - (analiza wrażliwości ze względu na cenę biletu)

Wzrost cen biletów oraz kosztów mandatów implikujący wzrost średniego przychodu jest zależnością, która zgadza się z naszymi oczekiwaniami. Brak wpływu liczby kontroli na średnie przychody, może być tłumaczony większymi kosztami kontroli, które będą balansowały się z przychodami, które zostaną uzyskane z mandatów.

5.2 Analiza wrażliwości ze względu na koszt kontroli

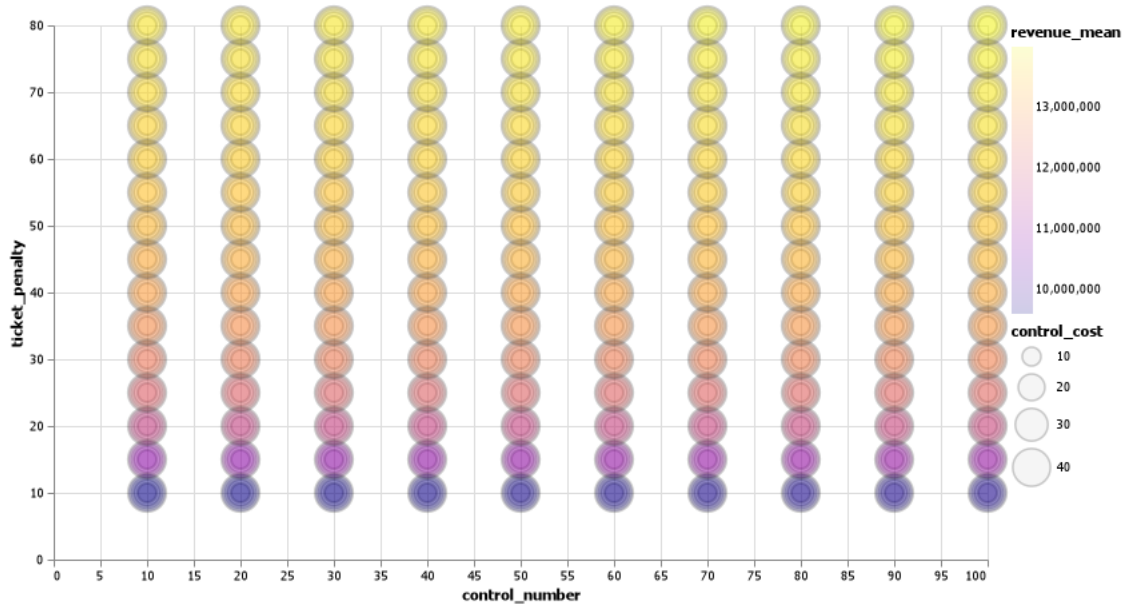
Jako drugą zmienną do analizy wrażliwości wybrano koszt kontroli pojedynczego kursu, w poniższej tabeli przedstawion optymalne rozwiązania dla poszczególnych scenariuszy. Podobnie jak przy cenie biletów rozważano zakres od dwukrotnego spadku kosztu do czterokrotnego wzrostu. Jak można zauważyć podane rozwiązania mają malejące wartości osiąganych przychodów względem kosztu kontroli.

Jak możemy zaobserwować w tabelce najwyższe średnie zarobki osiągane są dla kosztu kontroli równego 10 oraz kosztu mandatu równego 100.

Liczba kontroli	Koszt mandatu	Koszt kontroli	Średnie przychody
100	80.0	10.0	1.39907e7
100	80.0	20.0	1.39697e7
100	80.0	30.0	1.39355e7
100	80.0	40.0	1.39051e7

Tabela 2: Rozwiązania optymalne dla poszczególnych kosztów kontroli

Według symulacji zmiany z analizowanego przedziału nie mają wpływu na rozwiązanie optymalne, możliwe, że dla znacznie większych kosztów kontroli mogłaby zacząć spadać liczba zalecanych kontroli jednak na obecną chwilę analizowany przedział zawiera wszystkie prawdopodobne scenariusze i nie należy zmieniać rozwiązania w przypadku wahań.



Rysunek 6: Wykres bąbelkowy - (analiza wrażliwości ze względu na koszt kontroli)

6 Model regresji

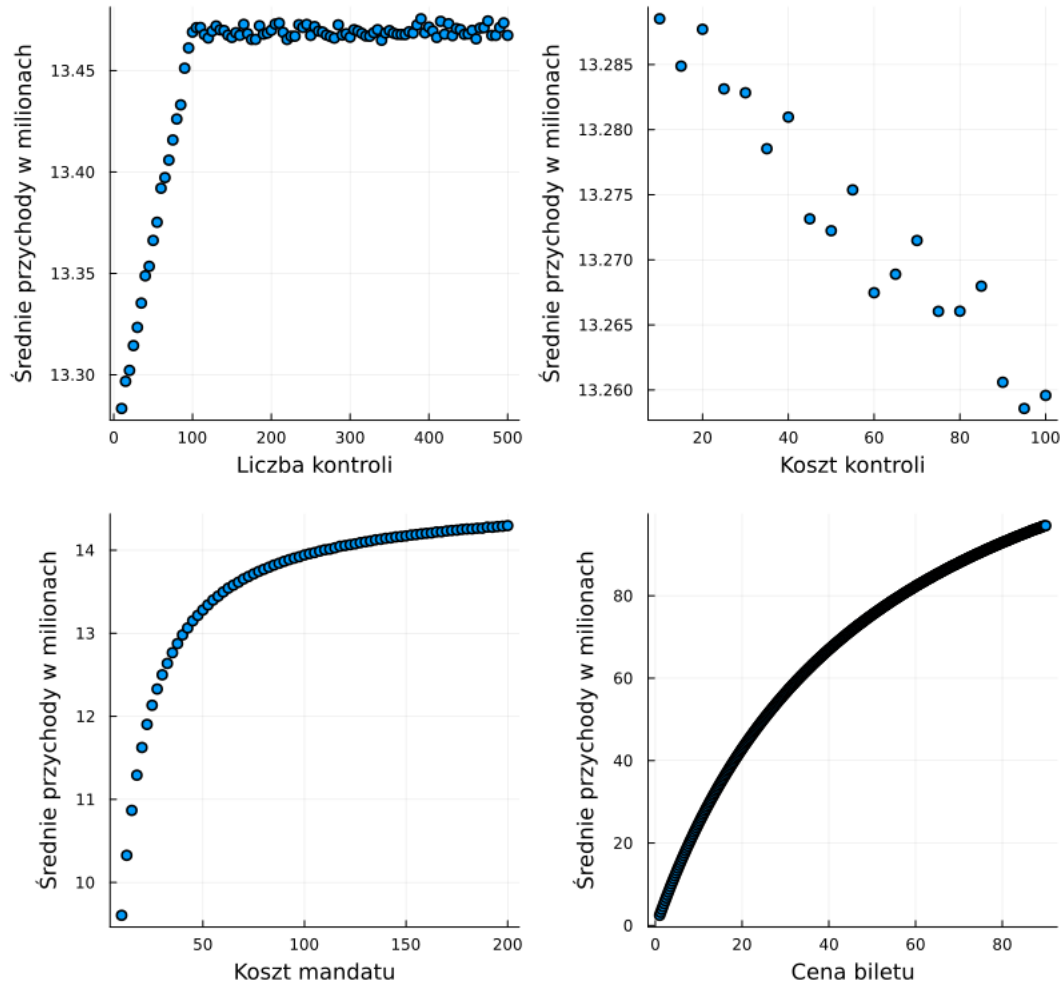
W celu lepszego zrozumienia jakie zmienne mają istotny wpływ na średnie przychody i jaka jest średnia wartość tego wpływu postanowiono skonstruować model regresji, który pozwolił odpowiedzieć na te pytania.

6.1 Badanie wpływu zmiennych na średnie przychody

W celu zbadania wpływu poszczególnych zmiennych na średnie przychody miesięczne komunikacji miejskiej zostały przeprowadzone symulacje z różnymi wartościami poszczególnych zmiennych.

Zmiany każdej ze zmiennych zostały zbadane osobno co pozwoliło na zaobserwowanie bezpośredniego wpływu - *ceteris paribus* poszczególnych zmiennych na średnie przychody jednostki.

Zmienne wpływające na średnie przychody



Rysunek 7

Na wykresie 7 można zaobserwować sposób w jaki kształtuje się zależność poszczególnych czynników od średnich przychodów jednostki. W szczególności można zaobserwować, że:

1. Wykres zależności średnich przychodów od liczby kontroli jest liniowy (patrzmy na wykres do liczby kontroli = 100)
2. Im mniejszy koszt kontroli tym większe średnie przychody - zależność liniowa
3. Średnie przychody w zależności od kosztu mandatu rosną logarytmicznie
4. Średnie przychody w zależności od ceny biletu rosną logarytmicznie

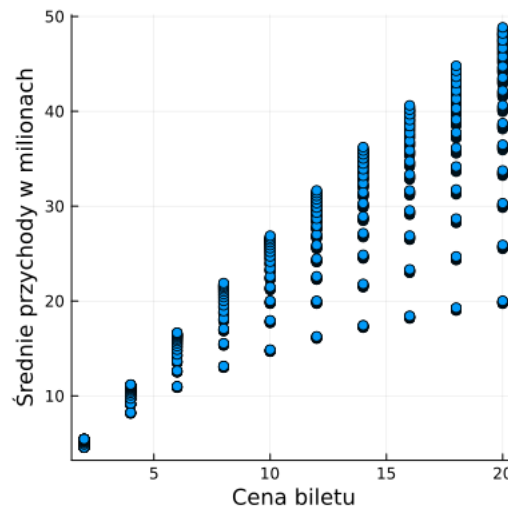
Dodatkowo należy dodać, że wykres ten powinien być zrobiony dla maksymalnej liczby kontroli 100. Zatem zależność ta jest liniowa. (W kodzie zostało już to uwzględnione)

6.2 Model regresji liniowej

Po wstępnym zbadaniu zależności zmiennych zbudowano model regresji liniowej w oparciu o zmienną: cena biletu. Zdecydowano się na taką zmienną z dwóch powodów. Po pierwsze jest to jedna ze zmiennych, dla których została przeprowadzona analiza wrażliwości i dzięki uwzględnieniu jej w modelu, możemy jeszcze lepiej pokazać w jaki sposób oddziałuje ona na średnie przychody. Po drugie zależność tej zmiennej od średnich przychodów dla małych wartości ceny biletu (do 100) kształtuje się w sposób liniowy.

$$\text{średni_przychód} = \text{cena_biletu} + \epsilon$$

Wykres 8 przedstawia zależność średnich przychodów od ceny biletu dla danych wysymulowanych na potrzeby regresji. Czyli dla danych, w których mamy różne wartości cen biletów, liczby kontroli oraz kosztów mandatu.



Rysunek 8

6.3 Wyniki modelu

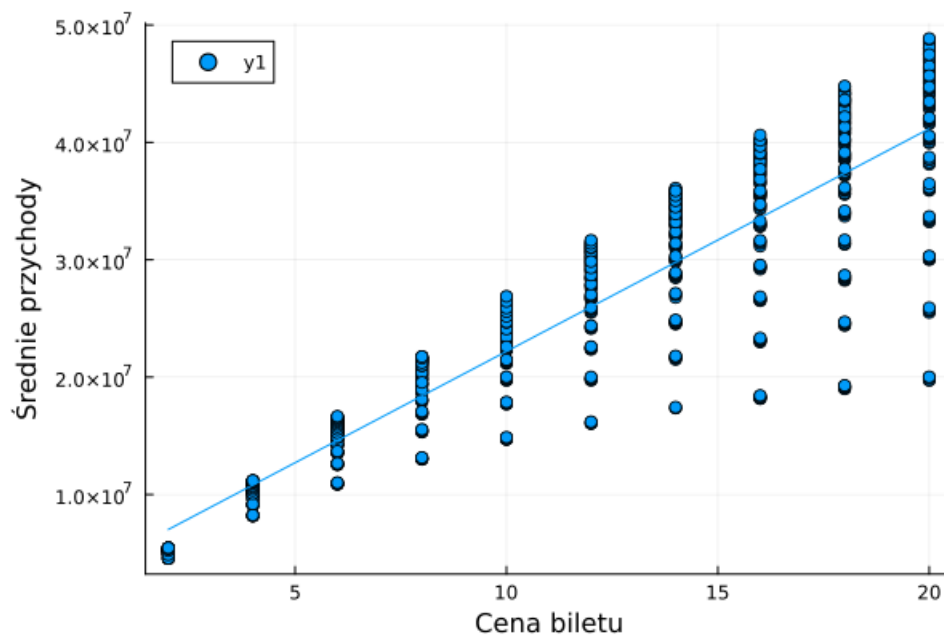
W celu oszacowania parametrów modelu nasz zbiór wysymulowanych danych został podzielony na zbiór testowy i treningowy w proporcjach odpowiednio: 30% i 70%.

Następnie parametry modelu zostały oszacowane na zbiorze danych treningowych. Otrzymano następujące wyniki:

Zmienne objaśniające	Oszacowania	P-value
Stała	3.22598	5.27352e-23
Cena biletu	1.89718	0.0

Tabela 3: Wyniki modelu regresji liniowej

Jak możemy zobaczyć w tabeli 3 zarówno stała jak i cena biletu są istotne statystycznie na poziomie istotności $\alpha = 0.01$. Warto podkreślić, że oszacowania w tabeli zostały podane w milionach. Zatem możemy powiedzieć, że wzrost ceny biletu o jednostkę powoduje średnio, *ceteris paribus*, wzrost średniego miesięcznego przychodu o około 1.89 miliona.



Rysunek 9

Kolejnym krokiem było przeprowadzenie prognoz na danych testowych oraz obliczenie statystyk mierzących błąd prognozy.

Miara błędu prognozy	MAE	MAPE	RMSE	MSE
Wartość miary	3.76622e6	0.183174	5.2902e6	2.79862e13

Tabela 4: Miary błędów prognozy

7 Wnioski i zalecenia

W trakcie analizy ustalono optymalne rozwiązanie jako 600 kontroli dziennych oraz maksymalny rozważany koszt kary za brak biletu czyli 80 PLN. Należy zauważyć, że w modelu przyjęto założenie o odwrotnej zależności prawdopodobieństwa jazdy bez biletu do wysokości kary, dlatego dla wyższych kar przewiduje się ogółem mniejszy odsetek gapowiczów, co może czynić rzadsze kontrole mniej skutecznymi. W analizie wrażliwości pochyłono się nad możliwymi zmianami ceny biletów i kosztu pojedynczej kontroli kursu. Rozważano scenariusze od ponad dwukrotnego spadku ceny biletu do czterokrotnej podwyżki, w przypadku kosztu kontroli zauważono spadek przychodów wraz ze wzrostem kosztu, jednak rozwiązania optymalne pozostały bez zmian. W ostatniej części raportu dokonano analizy regresji przychodów względem zmiennych podstawowych i tych z analizy wrażliwości. Jedynie koszt kontroli wpływał negatywnie na przychody, pozostałe zmienne miały dodatni wpływ jednak tylko dla liczby kontroli był on liniowy. Uzyskane wyniki wskazują, że znalezione rozwiązanie jest stabilne względem pozostałych czynników dla realnych wartości.

8 Bibliografia

1. Basiglio S., Mondinio M., Ritaccio D., Free Riders and public transport, 2012-2013

Spis rysunków

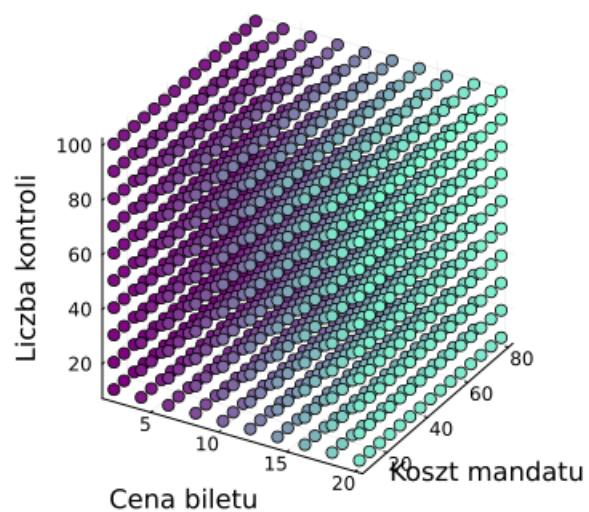
1	Dzienny przychód działalności transportowej w ciągu miesiąca	5
2	Średni zysk w zależności od mandatu i liczby kontroli	6
3	Rozwiązania pareto optymalne - liczba kontroli	6
4	Rozwiązania pareto optymalne - koszt mandatu	7
5	Wykres bąbelkowy - (analiza wrażliwości ze względu na cenę biletu)	8
6	Wykres bąbelkowy - (analiza wrażliwości ze względu na koszt kontroli)	9
7	10
8	11
9	12
10	15
11	15

Spis tabel

1	Rozwiązania optymalne dla poszczególnych cen biletów	7
2	Rozwiązania optymalne dla poszczególnych kosztów kontroli	9
3	Wyniki modelu regresji liniowej	12

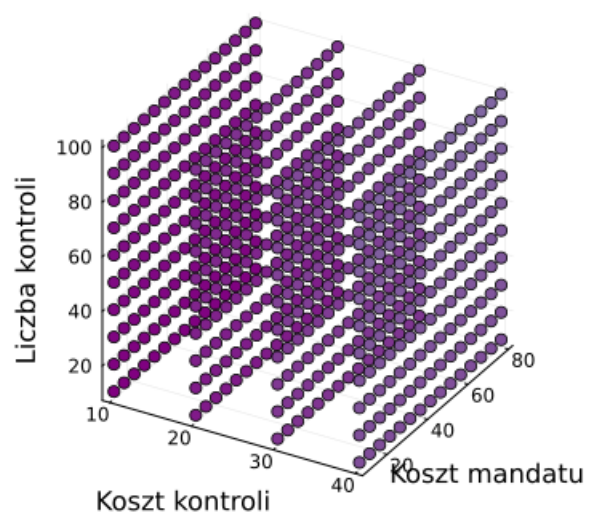
4	Miary błędów prognozy	12
---	---------------------------------	----

Wykres 3D - analiza wrażliwości ceny biletu



Rysunek 10

Wykres 3D - analiza wrażliwości kosztu kontroli



Rysunek 11