

Spaceship Titanic

Data Mining

Katsiaryna Bokhan

MiNI PW

30 października 2023

Plan prezentacji

① Wydawanie pieniędzy

- 1.1. Ciekawostka. Czy ludzie w hibernacji wydają pieniądze?
- 1.2. W jaki sposób wydawanie pieniędzy zależy od CabinDeck i statusu VIP?
- 1.3. W jaki sposób wydawanie pieniędzy na różne kategorie zależy od Wieku?

② Czy ludzie z różnych planet podróżują razem?

③ O jakich ludzi możemy powiedzieć, że zostali przetransportowani do innego wymiaru?

- 3.1. Jak CryoSleep i Transported są związane?
- 3.2. `CryoSleep = True`
- 3.3. `CryoSleep = False`
- 3.4. Osoby jakiego wieku najchętniej podróżują w hibernacji?
- 3.5. Czy transportation coefficient zależy od wieku podróżującego?

④ Uczenie Maszynowe

1.1. Czy ludzie w hibernacji wydają pieniądze?

```
In [6]: data_separated[data_separated['CryoSleep'] == True]['TotalSpending'].sum()
```

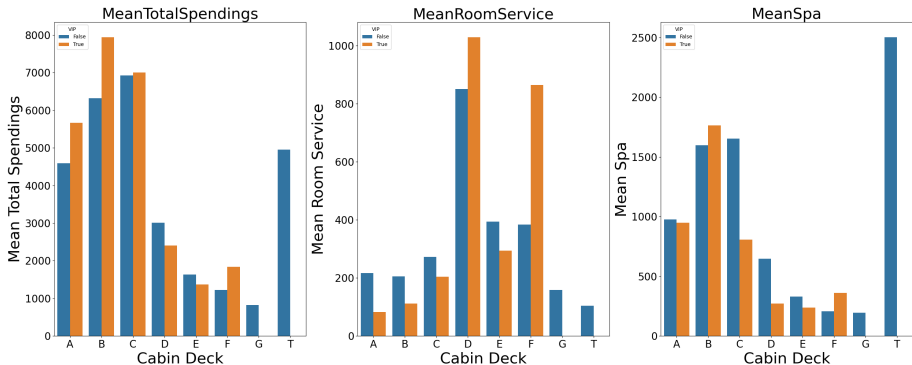
```
Out[6]: 0.0
```

Wniosek 1.1.

Ludzie w stanie hibernacji nie wydają pieniędzy.

Więc, kiedy będę liczyła średnie kwoty wydanych pieniędzy, to ludzi w hibernacji liczyć tam nie będę.

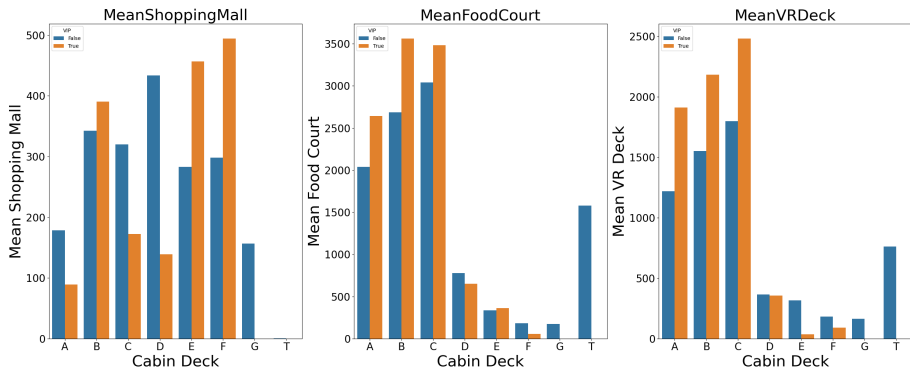
1.2. W jaki sposób wydawanie pieniędzy zależy od CabinDeck i statusu VIP?



Wniosek 1.2. (część 1)

- Ludzie z CabinDecków A, B, C, T średnio wydają znacznie więcej pieniędzy niż z CabinDecków D, E, F, G.
- W CabinDeckach G i T nie ma ludzi (nie w hibernacji) z VIP statusem.
- Ludzie w CabinDecku T wydają dużo na Spa, a z D i F na RoomService.

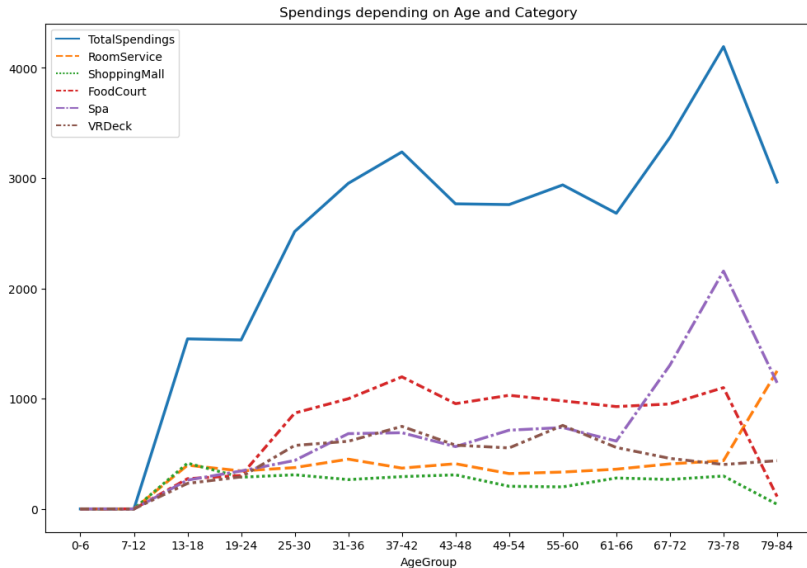
1.2. W jaki sposób wydawanie pieniędzy zależy od CabinDeck i statusu VIP?



Wniosek 1.2. (część 2)

- Ludzie z CabinDecków A, B, C średnio wydają znacznie więcej pieniędzy na VRDeck i FoodCourt, niż inni.

1.3. W jaki sposób wydawanie pieniędzy na różne kategorie zależy od Wiek?



1.3. W jaki sposób wydawanie pieniędzy na różne kategorie zależy od Wieku?

Wniosek 1.3.

- Ludzie, mające 12 lat i mniej nie wydają pieniędzy.
- Kwota wydanych w sumie pieniędzy rośnie wraz z wiekiem.
- Kwota wydanych pieniędzy na Spa rośnie wraz z wiekiem.
- Kwota wydanych pieniędzy na RoomService, ShoppingMall nie zależy od wieku osoby.
- Kwota wydanych pieniędzy na FoodCourt i VRDeck rośnie do 37-42 lat, potem nie zmienia się.

Plan prezentacji

① Wydawanie pieniędzy

- 1.1. Ciekawostka. Czy ludzie w hibernacji wydają pieniądze?
- 1.2. W jaki sposób wydawanie pieniędzy zależy od CabinDeck i statusu VIP?
- 1.3. W jaki sposób wydawanie pieniędzy na różne kategorie zależy od Wieku?

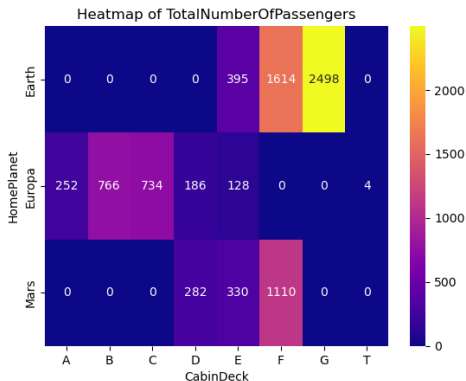
② Czy ludzie z różnych planet podróżują razem?

③ O jakich ludzi możemy powiedzieć, że zostali przetransportowani do innego wymiaru?

- 3.1. Jak CryoSleep i Transported są związane?
- 3.2. `CryoSleep = True`
- 3.3. `CryoSleep = False`
- 3.4. Osoby jakiego wieku najchętniej podróżują w hibernacji?
- 3.5. Czy transportation coefficient zależy od wieku podróżującego?

④ Uczenie Maszynowe

2. Czy ludzie z różnych planet podróżują razem?



Wniosek 2.

- W CabinDeck T są tylko 4 osoby (już wiemy, że uwielbiają spa :))
- Ludzie z Ziemi podróżują tylko na CabinDeckach E, F, G.
- Ludzie z Marsa - tylko na CabinDeckach D, E, F
- Ludzie z Europy - głównie w CabinDeckach B, C (następnie A, D, E, T)

Plan prezentacji

① Wydawanie pieniędzy

- 1.1. Ciekawostka. Czy ludzie w hibernacji wydają pieniądze?
- 1.2. W jaki sposób wydawanie pieniędzy zależy od CabinDeck i statusu VIP?
- 1.3. W jaki sposób wydawanie pieniędzy na różne kategorie zależy od Wieku?

② Czy ludzie z różnych planet podróżują razem?

③ O jakich ludzi możemy powiedzieć, że zostali przetransportowani do innego wymiaru?

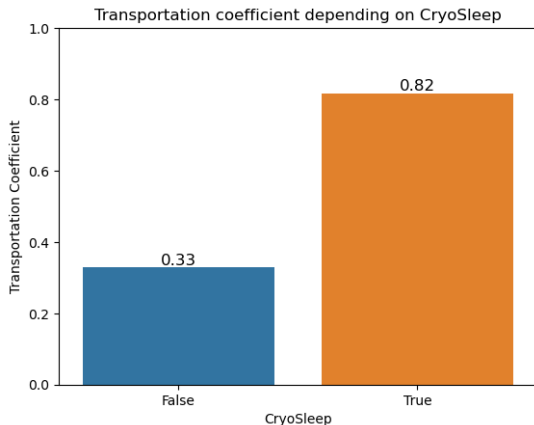
- 3.1. Jak CryoSleep i Transported są związane?
- 3.2. `CryoSleep = True`
- 3.3. `CryoSleep = False`
- 3.4. Osoby jakiego wieku najchętniej podróżują w hibernacji?
- 3.5. Czy transportation coefficient zależy od wieku podróżującego?

④ Uczenie Maszynowe

3.1. Jak CryoSleep i Transported są związane?

Definicja

Transportation coefficient ($\text{CryoSleep} == \text{True}$) =
$$\frac{\text{liczba osób z CryoSleep} == \text{True} \ \& \ \text{Transported} == \text{TRUE}}{\text{ogólna liczba osób na statku z CryoSleep} == \text{True}}$$



3.1. Jak CryoSleep i Transported są związane?

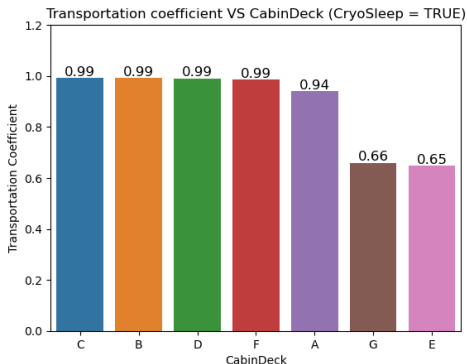
Wniosek 3.1.

- 82 % ludzi, którzy byli w hibernacji zostali przetransportowani do innego wymiaru.

Widzimy, że CryoSleep to bardzo ważny parameter, na podstawie którego już możemy dość dużo powiedzieć. Spróbujmy znaleźć jeszcze inne takie bardzo ważne parametry.

Rozważamy teraz tylko osób w
hibernacji

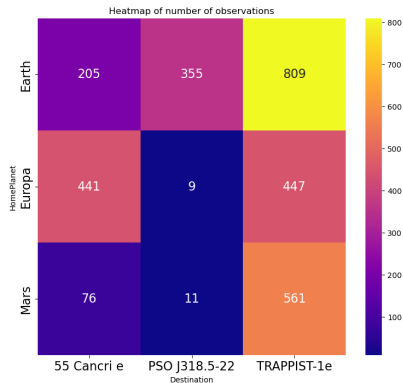
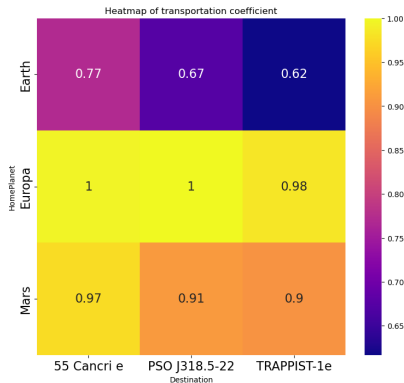
3.2.



Wniosek 3.1.1.

- 99,3197% osób na pokładzie C zostało przetransportowano (292 z 294).
- 99,2840% osób na pokładzie B, zostało przetransportowano (416 z 419).
- 99,0385% osób na pokładzie D zostało przetransportowano (103 z 104).
- 98,5455% osób na pokładzie F zostało przetransportowano (542 z 550).

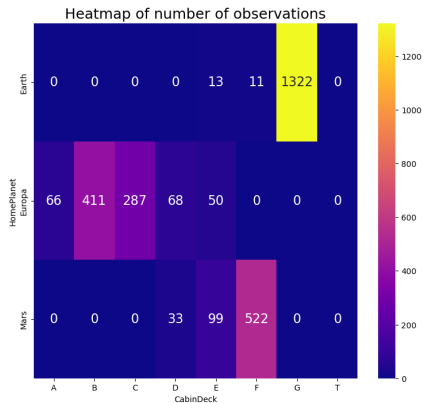
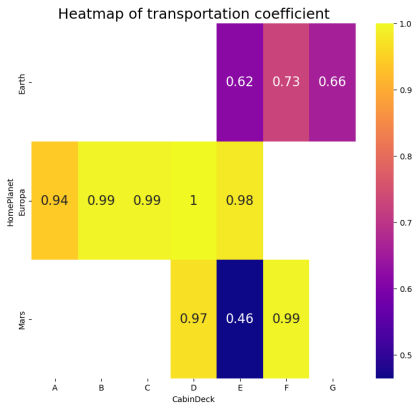
3.2.(c.d.)



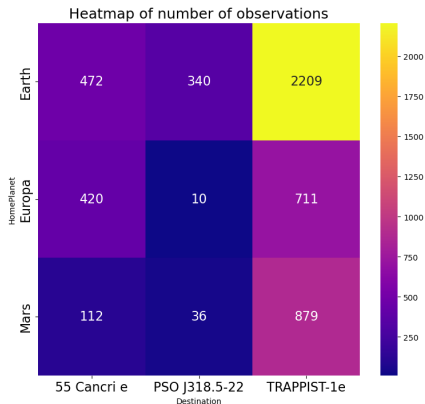
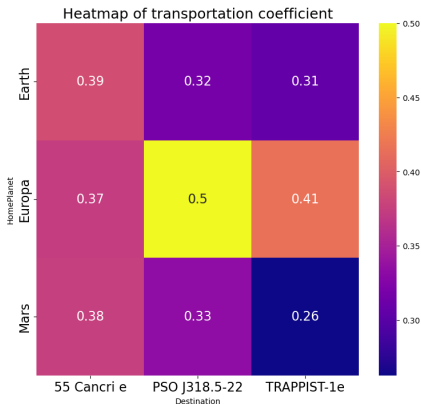
Wniosek 3.1.4.

- 99,5465% Europa – > 55 Cancri e zostało przetransportowane.
- 100% Europa – > PSO J318.5-22 zostało przetransportowane.
- 98,2103% Europa – > TRAPPIST-1e, zostało przetransportowane.
- 97,3684% Mars – > PSO J318.5-22 , zostało przetransportowane.

3.2.(c.d.)



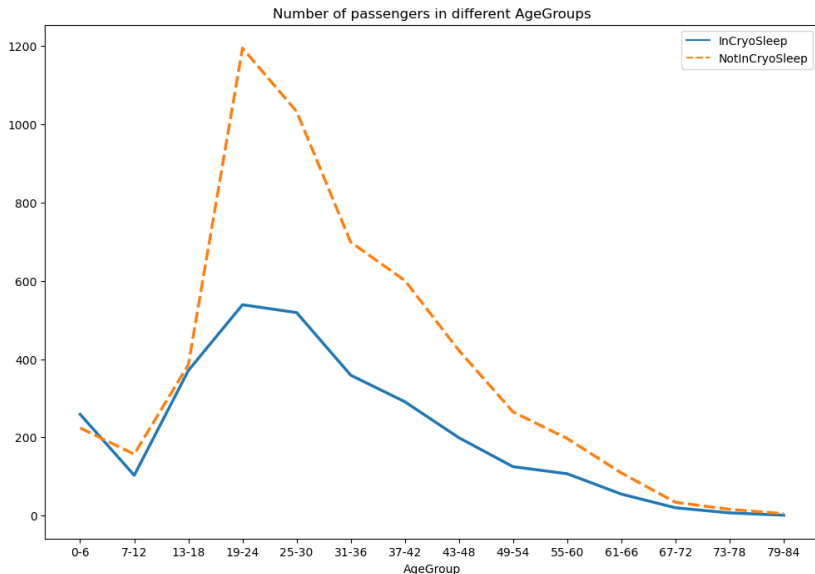
Rozważamy teraz osób nie w hibernacji. Najbardziej ciekawy wykres według mnie.



Wniosek 3.2.

- Najmniejszy transportation coefficient - Mars \rightarrow TRAPPIST-1e. - około 26 %
- Najmniej osób zostało przetransportowano pochodzących c Marsa

3.4. Osoby jakiego wieku najchętniej podróżują w hibernacji?

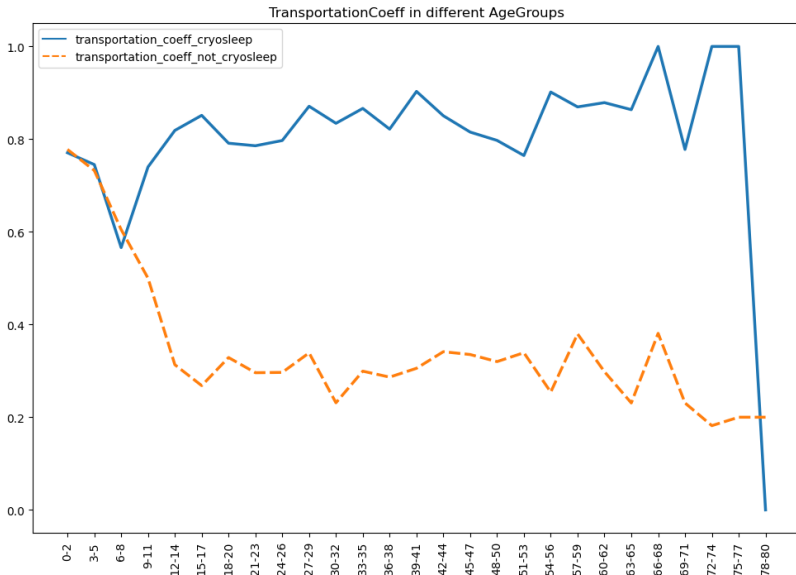


3.4. Osoby jakiego wieku najchętniej podróżują w hibernacji?

Wniosek 3.4.

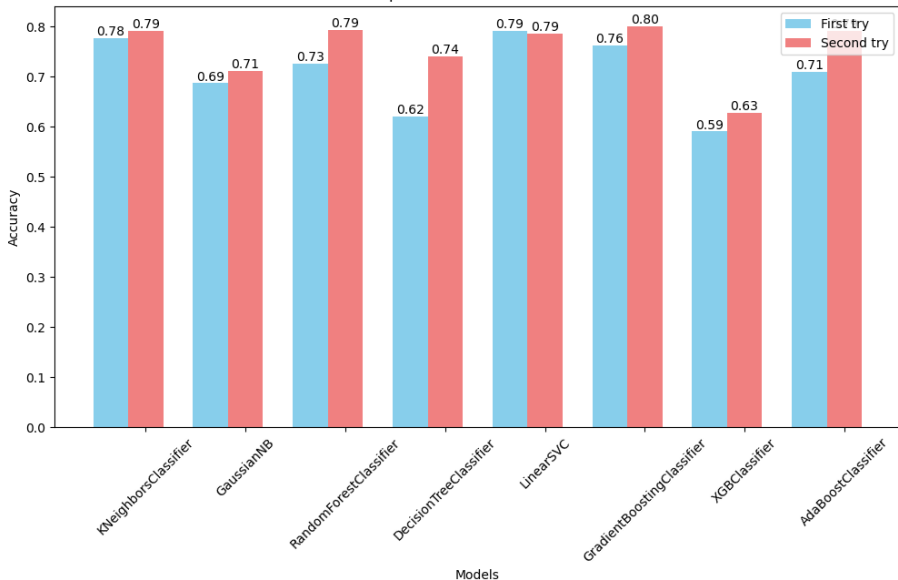
- Osoby poniżej 18. roku życia i osoby powyżej 67 roku życia wybierają korzystanie z hibernacji z taką samą częstotliwością, jak niebycie w hibernacji. Inni preferują być nie w hibernacji.
- Największa różnica między osobami, które wolą podróżować w hibernacji lub nie, jest widoczna w grupie wiekowej 19-24 lata

3.5. Czy transportation coefficient zależy od wieku podróżującego? Prawie nie.



Wyniki z użycia algorytmów uczenia maczynowego

Comparison of Model Accuracies



Jakie były zmiany?

Nowe wartości parametrów

$Spa = Spa / TotalSpending$

i analogiczne dla innych kategorii wydania pieniędzy.

Jakie narzędzia były używane podczas data mining?

- Jupyter Notebook
- R (dplyr), Python(pandas)
- Seaborn, Matplotlib - visualizacja
- Scikit-learn - uczenie maszynowe

The End