

TuristiSažimanje

August 1, 2025

```
[1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[2]: turisti = pd.read_excel("../..//datasets/Turisti.xlsx")
```

```
[3]: turisti.head()
```

```
[3]:
```

	Country	City	NoDays	Motivation	Price	NrPeople	NrCh	Service	\
0	Italija	Padova	10.0	tourism	1700.0	4.0	2.0	700.0	
1	Portugal	Lisbon	11.0	tourism	2475.0	6.0	4.0	250.0	
2	Italija	Milano	5.0	tourism	850.0	5.0	4.0	250.0	
3	Španjolska	Barcelona	7.0	tourism	1505.0	5.0	3.0	500.0	
4	Njemačka	Berlin	12.0	tourism	2280.0	4.0	2.0	700.0	

	Transport	Loyalty	Paymenttype
0	bus	6.0	2.0
1	train	3.0	2.0
2	airplane	4.0	1.0
3	bus	4.0	1.0
4	train	5.0	2.0

```
[4]: turisti.describe()
```

```
[4]:
```

	NoDays	Price	NrPeople	NrCh	Service	\
count	299.000000	299.000000	298.000000	298.000000	299.000000	
mean	8.187291	2228.377926	3.174497	2.194631	794.481605	
std	2.261074	9749.828055	1.498513	12.831557	224.680195	
min	5.000000	85.000000	1.000000	0.000000	250.000000	
25%	7.000000	1190.000000	2.000000	0.000000	700.000000	
50%	9.000000	1700.000000	4.000000	2.000000	700.000000	
75%	10.000000	2070.000000	4.000000	2.000000	1000.000000	
max	12.000000	170000.000000	6.000000	222.000000	1000.000000	

	Loyalty	Paymenttype
count	299.000000	299.000000
mean	3.608696	1.936455
std	1.746730	0.785337

min	1.000000	1.000000
25%	2.000000	1.000000
50%	4.000000	2.000000
75%	5.000000	3.000000
max	6.000000	3.000000

```
[5]: turisti.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300 entries, 0 to 299
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Country         299 non-null   object
1   City            298 non-null   object
2   NoDays          299 non-null   float64
3   Motivation      297 non-null   object
4   Price           299 non-null   float64
5   NrPeople        298 non-null   float64
6   NrCh            298 non-null   float64
7   Service         299 non-null   float64
8   Transport       298 non-null   object
9   Loyalty         299 non-null   float64
10  Paymenttype     299 non-null   float64
dtypes: float64(7), object(4)
memory usage: 25.9+ KB
```

```
[6]: turisti.Country.value_counts()
```

```
[6]: Country
Italija      69
Španjolska  61
Francuska   59
Njemačka    46
Portugal    38
Švedska     16
Island      10
Name: count, dtype: int64
```

```
[7]: turisti.City.value_counts()
```

```
[7]: City
Lisbon      38
Barcelona   32
Paris       32
Madrid      29
Berlin      27
Lyon        27
```

```
Padova      24
Rome        23
Milano      20
Frankfurt   19
Stockholm   16
Reykjavík   10
rome         1
Name: count, dtype: int64
```

```
[8]: turisti.Motivation.value_counts()
```

```
[8]: Motivation
tourism      237
business      59
busines        1
Name: count, dtype: int64
```

```
[9]: turisti.Transport.value_counts()
```

```
[9]: Transport
bus          122
train         85
airplane      68
Airplane      23
Name: count, dtype: int64
```

```
[10]: turisti.corr(numeric_only = True)
```

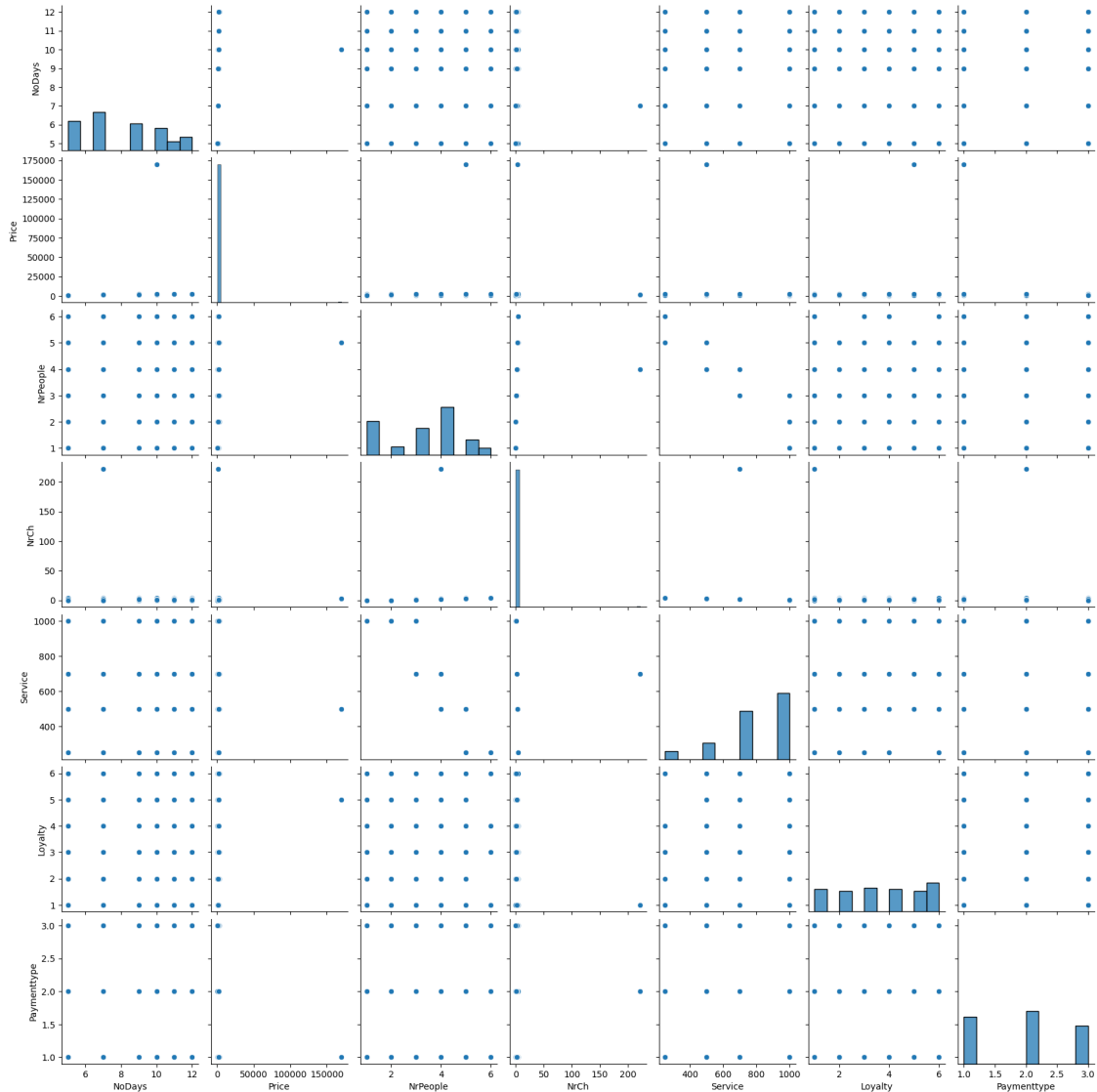
```
[10]:
```

	NoDays	Price	NrPeople	NrCh	Service	Loyalty \
NoDays	1.000000	0.097999	0.332548	-0.005752	-0.202398	0.009272
Price	0.097999	1.000000	0.089489	0.003674	-0.087627	0.047398
NrPeople	0.332548	0.089489	1.000000	0.121863	-0.886605	-0.062233
NrCh	-0.005752	0.003674	0.121863	1.000000	-0.113301	-0.091556
Service	-0.202398	-0.087627	-0.886605	-0.113301	1.000000	0.056898
Loyalty	0.009272	0.047398	-0.062233	-0.091556	0.056898	1.000000
Paymenttype	-0.014063	-0.070582	0.074845	0.012575	-0.098985	0.016060

	Paymenttype
NoDays	-0.014063
Price	-0.070582
NrPeople	0.074845
NrCh	0.012575
Service	-0.098985
Loyalty	0.016060
Paymenttype	1.000000

```
[11]: sns.pairplot (turisti)
```

[11]: <seaborn.axisgrid.PairGrid at 0x7ff8d49b9390>



[12]: turisti.sort_values (by = ["Price"])

```
[12]:
```

	Country	City	NoDays	Motivation	Price	NrPeople	NrCh	Service	\
240	Italija	Milano	5.0	business	85.0	1.0	0.0	1000.0	
295	Italija	Rome	5.0	business	850.0	1.0	0.0	1000.0	
22	Italija	Padova	5.0	tourism	850.0	5.0	3.0	500.0	
2	Italija	Milano	5.0	tourism	850.0	5.0	4.0	250.0	
34	Italija	Rome	5.0	tourism	850.0	4.0	2.0	700.0	
..	...	...	...	...	...	...	...	...	
190	Švedska	Stockholm	12.0	tourism	2760.0	5.0	3.0	500.0	
178	Švedska	Stockholm	12.0	tourism	2760.0	3.0	1.0	1000.0	

231	Švedska	Stockholm	12.0	tourism	2760.0	2.0	0.0	1000.0
82	Italija	Padova	10.0	tourism	170000.0	5.0	3.0	500.0
49	Italija	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	Transport	Loyalty	Paymenttype
240	airplane	4.0	2.0
295	airplane	2.0	2.0
22	airplane	4.0	3.0
2	airplane	4.0	1.0
34	airplane	5.0	1.0
..	...	...	...
190	Airplane	6.0	3.0
178	Airplane	2.0	1.0
231	Airplane	3.0	1.0
82	bus	5.0	1.0
49	NaN	NaN	NaN

[300 rows x 11 columns]

```
[13]: outlier_index_Price = [240,82]
```

```
[14]: mean_non_outliers = turisti.drop(outlier_index_Price)["Price"].mean()
```

```
[15]: mean_non_outliers
```

```
[15]: np.float64(1670.7070707070707)
```

```
[16]: turisti.loc[outlier_index_Price, "Price"] = mean_non_outliers
```

```
[17]: turisti.describe()
```

```
[17]:
```

	NoDays	Price	NrPeople	NrCh	Service \
count	299.000000	299.000000	298.000000	298.000000	299.000000
mean	8.187291	1670.707071	3.174497	2.194631	794.481605
std	2.261074	528.860214	1.498513	12.831557	224.680195
min	5.000000	850.000000	1.000000	0.000000	250.000000
25%	7.000000	1190.000000	2.000000	0.000000	700.000000
50%	9.000000	1700.000000	4.000000	2.000000	700.000000
75%	10.000000	2070.000000	4.000000	2.000000	1000.000000
max	12.000000	2760.000000	6.000000	222.000000	1000.000000

	Loyalty	Paymenttype
count	299.000000	299.000000
mean	3.608696	1.936455
std	1.746730	0.785337
min	1.000000	1.000000
25%	2.000000	1.000000

50%	4.000000	2.000000
75%	5.000000	3.000000
max	6.000000	3.000000

```
[18]: turisti.sort_values (by = ["NrCh"])
```

```
[18]:
```

	Country	City	NoDays	Motivation	Price	NrPeople	NrCh	\
295	Italija	Rome	5.0	business	850.0	1.0	0.0	
21	Španjolska	Barcelona	5.0	business	1075.0	1.0	0.0	
24	Francuska	Paris	7.0	business	1435.0	1.0	0.0	
272	Španjolska	Barcelona	5.0	business	1075.0	1.0	0.0	
265	Španjolska	Madrid	10.0	tourism	2150.0	1.0	0.0	
..	...	...	...	...	...	...	...	
27	Njemačka	Frankfurt	9.0	tourism	1710.0	6.0	4.0	
1	Portugal	Lisbon	11.0	tourism	2475.0	6.0	4.0	
57	Francuska	Paris	7.0	tourism	1435.0	4.0	222.0	
49	Italija	NaN	NaN	NaN	NaN	NaN	NaN	
241	Italija	Milano	5.0	business	850.0	1.0	NaN	

	Service	Transport	Loyalty	Paymenttype
295	1000.0	airplane	2.0	2.0
21	1000.0	airplane	6.0	1.0
24	1000.0	bus	2.0	3.0
272	1000.0	airplane	4.0	1.0
265	1000.0	bus	3.0	2.0
..	...	...	...	...
27	250.0	train	3.0	2.0
1	250.0	train	3.0	2.0
57	700.0	bus	1.0	2.0
49	NaN	NaN	NaN	NaN
241	1000.0	airplane	2.0	2.0

[300 rows x 11 columns]

```
[19]: outlier_index_NrCh = [57]
```

```
[20]: mean_non_outliers_NrCh = turisti.drop(outlier_index_NrCh) ["NrCh"].mean()
```

```
[21]: mean_non_outliers_NrCh
```

```
[21]: np.float64(1.4545454545454546)
```

```
[22]: mean_non_outliers_NrCh
```

```
[22]: np.float64(1.4545454545454546)
```

```
[23]: mean_non_outliers_NrCh = round(mean_non_outliers_NrCh)
```

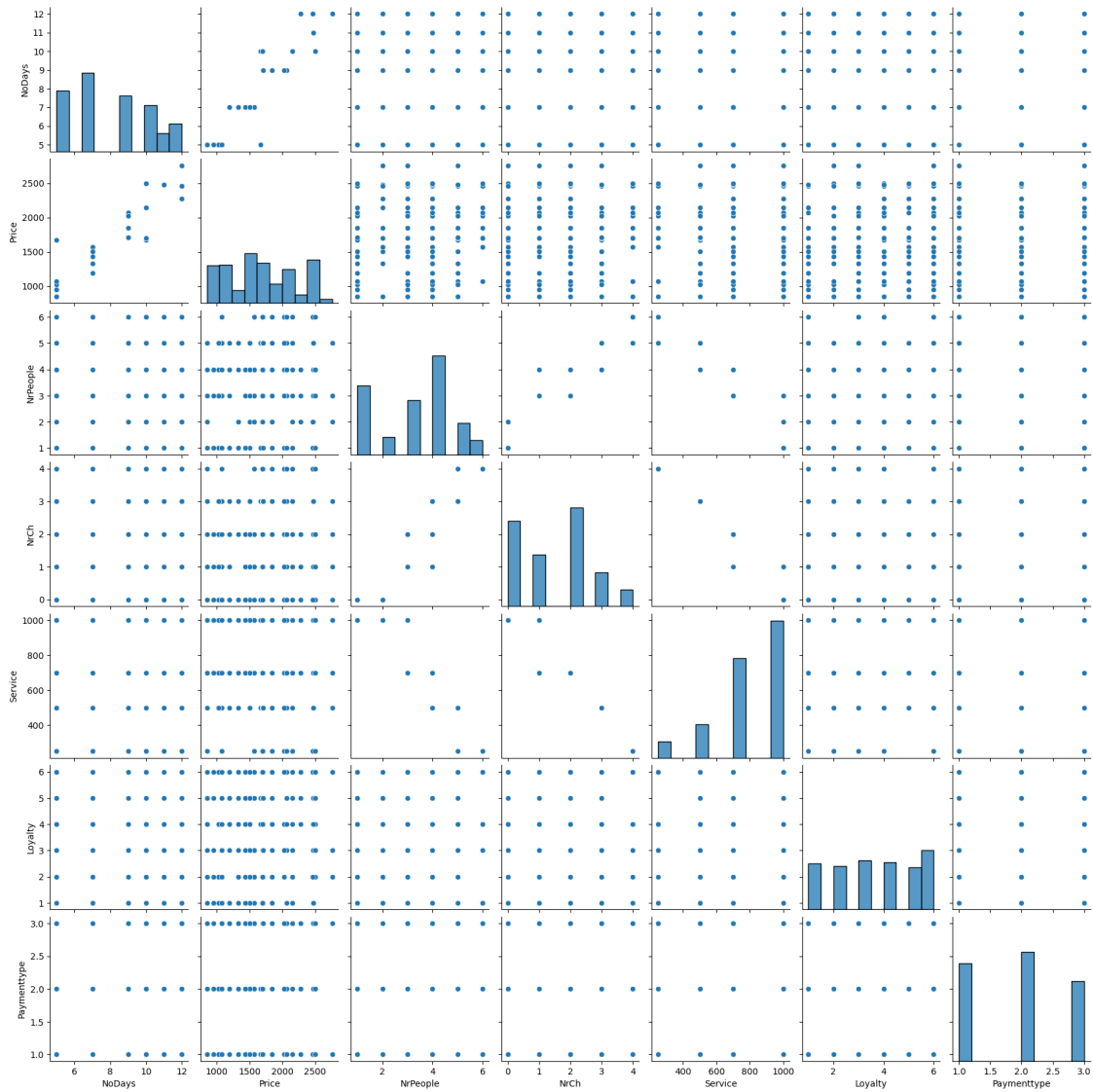
```
[24]: mean_non_outliers_NrCh
```

```
[24]: 1
```

```
[25]: turisti.loc[outlier_index_NrCh, "NrCh"] = mean_non_outliers_NrCh
```

```
[26]: sns.pairplot (turisti)
```

```
[26]: <seaborn.axisgrid.PairGrid at 0x7ff8d49c2b10>
```



```
[27]: turisti.corr(numeric_only=True)
```

```
[27]:
```

	NoDays	Price	NrPeople	NrCh	Service	Loyalty \
NoDays	1.000000	0.936196	0.332548	0.268794	-0.202398	0.009272
Price	0.936196	1.000000	0.331859	0.269210	-0.206346	0.025314
NrPeople	0.332548	0.331859	1.000000	0.967994	-0.886605	-0.062233
NrCh	0.268794	0.269210	0.967994	1.000000	-0.957040	-0.051465
Service	-0.202398	-0.206346	-0.886605	-0.957040	1.000000	0.056898
Loyalty	0.009272	0.025314	-0.062233	-0.051465	0.056898	1.000000
Paymenttype	-0.014063	-0.026802	0.074845	0.084584	-0.098985	0.016060

	Paymenttype
NoDays	-0.014063
Price	-0.026802
NrPeople	0.074845
NrCh	0.084584
Service	-0.098985
Loyalty	0.016060
Paymenttype	1.000000

```
[28]: turisti["City"] = turisti["City"].replace({"rome": "Rome"})
```

```
[29]: turisti.City.value_counts()
```

```
[29]: City
Lisbon      38
Barcelona   32
Paris       32
Madrid      29
Berlin      27
Lyon        27
Rome        24
Padova      24
Milano      20
Frankfurt   19
Stockholm   16
Reykjavík   10
Name: count, dtype: int64
```

```
[30]: turisti["Motivation"] = turisti["Motivation"].replace({"busines": "business"})
```

```
[31]: turisti.Motivation.value_counts()
```

```
[31]: Motivation
tourism      237
business      60
Name: count, dtype: int64
```

```
[32]: turisti["Transport"] = turisti["Transport"].replace({"Airplane": "airplane"})
```

```
[33]: turisti.Transport.value_counts()
```

```
[33]: Transport
      bus      122
      airplane  91
      train    85
      Name: count, dtype: int64
```

```
[34]: isna_values = turisti[turisti.isna().any(axis=1)]
```

```
[35]: isna_values
```

```
[35]:
```

	Country	City	NoDays	Motivation	Price	NrPeople	NrCh	\
39	NaN	Madrid	5.0	tourism	1075.0	4.0	2.0	
49	Italija	NaN	NaN	NaN	NaN	NaN	NaN	
83	Španjolska	Madrid	7.0	NaN	1505.0	4.0	2.0	
238	Island	Reykjavík	10.0	tourism	2500.0	1.0	0.0	
241	Italija	Milano	5.0	business	850.0	1.0	NaN	
270	Španjolska	NaN	7.0	NaN	1505.0	NaN	2.0	

	Service	Transport	Loyalty	Paymenttype
39	700.0	airplane	4.0	1.0
49	NaN	NaN	NaN	NaN
83	700.0	bus	4.0	2.0
238	1000.0	NaN	6.0	2.0
241	1000.0	airplane	2.0	2.0
270	700.0	bus	5.0	1.0

```
[36]: turisti.dropna(thresh = 5, inplace = True)
```

```
[37]: prosjek_People = turisti["NrPeople"].mean()
```

```
[38]: prosjek_People
```

```
[38]: np.float64(3.174496644295302)
```

```
[39]: prosjek_People = round (prosjek_People)
```

```
[40]: turisti["NrPeople"] = turisti["NrPeople"].fillna (prosjek_People)
```

```
[41]: isna_values = turisti[turisti.isna().any(axis=1)]
```

```
[42]: isna_values
```

```
[42]:
```

	Country	City	NoDays	Motivation	Price	NrPeople	NrCh	\
39	NaN	Madrid	5.0	tourism	1075.0	4.0	2.0	
83	Španjolska	Madrid	7.0	NaN	1505.0	4.0	2.0	
238	Island	Reykjavík	10.0	tourism	2500.0	1.0	0.0	

241	Italija	Milano	5.0	business	850.0	1.0	NaN
270	Španjolska	NaN	7.0	NaN	1505.0	3.0	2.0

	Service	Transport	Loyalty	Paymenttype
39	700.0	airplane	4.0	1.0
83	700.0	bus	4.0	2.0
238	1000.0	NaN	6.0	2.0
241	1000.0	airplane	2.0	2.0
270	700.0	bus	5.0	1.0

```
[43]: mean_non_outliers_NrCh
```

```
[43]: 1
```

```
[44]: turisti["NrCh"] = turisti["NrCh"].fillna(mean_non_outliers_NrCh)
```

```
[45]: isna_values = turisti[turisti.isna().any(axis=1)]
```

```
[46]: isna_values
```

```
[46]:
```

	Country	City	NoDays	Motivation	Price	NrPeople	NrCh	\
39	NaN	Madrid	5.0	tourism	1075.0	4.0	2.0	
83	Španjolska	Madrid	7.0	NaN	1505.0	4.0	2.0	
238	Island	Reykjavík	10.0	tourism	2500.0	1.0	0.0	
270	Španjolska	NaN	7.0	NaN	1505.0	3.0	2.0	

	Service	Transport	Loyalty	Paymenttype
39	700.0	airplane	4.0	1.0
83	700.0	bus	4.0	2.0
238	1000.0	NaN	6.0	2.0
270	700.0	bus	5.0	1.0

```
[47]: turisti["Motivation"].mode()
```

```
[47]: 0    tourism
      Name: Motivation, dtype: object
```

```
[48]: turisti["Motivation"] = turisti["Motivation"].fillna(value = "tourism")
```

```
[49]: isna_values = turisti[turisti.isna().any(axis=1)]
```

```
[50]: turisti["Country"] = turisti["Country"].fillna(value = "Španjolska")
```

```
[51]: isna_values = turisti[turisti.isna().any(axis=1)]
```

```
[52]: turisti["Transport"].mode()
```

```
[52]: 0    bus
      Name: Transport, dtype: object
```

```
[53]: turisti["Transport"] = turisti["Transport"].fillna(value = "bus")
```

```
[54]: isna_values = turisti[turisti.isna().any(axis=1)]
```

```
[55]: isna_values
```

```
[55]:      Country City  NoDays Motivation  Price  NrPeople  NrCh  Service \
270  Španjolska  NaN     7.0    tourism  1505.0        3.0   2.0   700.0

      Transport  Loyalty  Paymenttype
270         bus     5.0           1.0
```

```
[56]: turisti.duplicated().sum()
```

```
[56]: np.int64(15)
```

```
[57]: turisti.drop_duplicates(inplace=True)
```

```
[58]: turisti
```

```
[58]:      Country      City  NoDays Motivation  Price  NrPeople  NrCh  \
0      Italija    Padova    10.0    tourism  1700.0        4.0   2.0
1      Portugal    Lisbon    11.0    tourism  2475.0        6.0   4.0
2      Italija    Milano     5.0    tourism   850.0        5.0   4.0
3  Španjolska  Barcelona     7.0    tourism  1505.0        5.0   3.0
4    Njemačka    Berlin    12.0    tourism  2280.0        4.0   2.0
..      ...      ...      ...      ...      ...      ...
295   Italija      Rome     5.0  business   850.0        1.0   0.0
296   Portugal    Lisbon    11.0    tourism  2475.0        3.0   1.0
297   Italija    Padova    10.0    tourism  1700.0        3.0   1.0
298    Island  Reykjavík    10.0    tourism  2500.0        4.0   2.0
299    Island  Reykjavík    10.0    tourism  2500.0        3.0   1.0

      Service Transport  Loyalty  Paymenttype
0      700.0        bus     6.0           2.0
1      250.0       train     3.0           2.0
2      250.0  airplane     4.0           1.0
3      500.0        bus     4.0           1.0
4      700.0       train     5.0           2.0
..      ...      ...      ...      ...
295    1000.0  airplane     2.0           2.0
296    1000.0       train     4.0           2.0
297    1000.0        bus     2.0           1.0
298     700.0  airplane     5.0           1.0
299    1000.0  airplane     6.0           2.0
```

[284 rows x 11 columns]