

Motivation

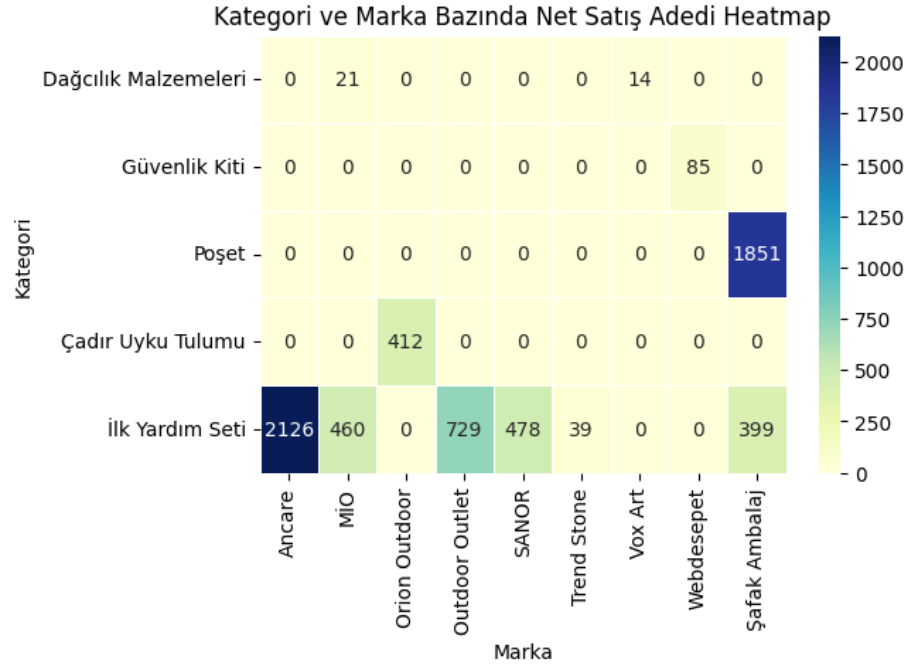
My primary motivation is analyzing my current sales on e-commerce. I will use my Trendyol sales data available in the Excel file, featuring columns such as Barcode, Product Name, Model Code, Category, Brand, Color, Size, Gross Sales Quantity, Cancelled Quantity, Cancel Rate, Returned Quantity, Return Rate, Net Sales Quantity, Gross Revenue, Net Revenue, Total Commission Amount, Average Commission Amount, Average Commission Rate, Average Selling Price, Current Selling Price, and Current Stock. By delving into these datasets, I aim to uncover patterns, trends, and correlations that can inform decision-making processes. The analysis will not only provide a comprehensive overview of sales performance but also contribute to optimizing pricing strategies and understanding customer preferences. So, this will be useful for decision-making and enhance overall business efficiency within the context of Trendyol's e-commerce platform.

Data Exploration

1-) I analyzed the provided data to gain insights into the sales performance of a product, focusing on key metrics such as sales quantity, cancellation rates, return rates, revenue figures, and commission-related information. This examination aims to assess the overall sales dynamics, identify potential areas for improvement, and understand the financial implications of the sales transactions.

Metric Name	Count	Mean	Std Dev	Min	25%	50%	75%	Max
Brut Sales Quantity	16	309.44	512.66	0.00	34.75	120.00	429.00	2085.00
Cancelled Quantity	16	9.06	19.42	0.00	0.00	1.00	4.25	66.00
Cancel Rate	16	2.16%	3.57%	0.00%	0.00%	1.01%	2.51%	14.38%
Return Quantity	16	3.19	5.15	0.00	0.00	0.50	4.25	15.00
Return Rate	16	1.99%	5.67%	0.00%	0.00%	0.21%	0.97%	23.00%
Net Sales Quantity	16	413.38	621.98	2.00	71.25	145.50	428.50	2021.00
Gross Revenue	16	9075.94	13690.85	281.00	2313.00	4177.00	9416.00	57317.00
Net Revenue	16	8650.72	13252.47	200.53	2188.25	4001.50	9132.00	55189.00
Total Commission Amount	16	1161.80	1954.55	8.72	321.49	550.27	1135.00	8143.03
Avg Commission Amount	16	18.28	49.48	1.43	3.04	4.11	6.45	202.39
Avg Commission Rate	16	16.94%	15.36%	7.29%	12.17%	14.21%	14.67%	73.95%
Avg Sales Price	16	159.62	464.58	19.05	24.63	28.35	45.41	1895.00
Current Sales Price	16	230.93	501.46	24.90	25.60	37.99	119.49	1895.00

2-) I sorted the data by net sales quantity, categorizing it based on product category and brand, with the objective of identifying the top-performing combination. This analysis allows us to pinpoint the category and brand that contribute the most to overall sales, providing valuable insights for strategic decision-making and potential optimization efforts within the product portfolio. Then I used a heatmap.



- Marka: Ancare, Net Satış Adedi: 2021
- Marka: Şafak Ambalaj, Net Satış Adedi: 1851
- Marka: Outdoor Outlet, Net Satış Adedi: 600
- Marka: SANOR, Net Satış Adedi: 478
- Marka: Orion Outdoor, Net Satış Adedi: 412
- Marka: MiO, Net Satış Adedi: 378
- Marka: Şafak Ambalaj, Net Satış Adedi: 235
- Marka: Şafak Ambalaj, Net Satış Adedi: 162
- Marka: Outdoor Outlet, Net Satış Adedi: 129
- Marka: Ancare, Net Satış Adedi: 105
- Marka: Webdesepet, Net Satış Adedi: 85
- Marka: MiO, Net Satış Adedi: 82
- Marka: Trend Stone, Net Satış Adedi: 39
- Marka: MiO, Net Satış Adedi: 21
- Marka: Vox Art, Net Satış Adedi: 14
- Marka: Şafak Ambalaj, Net Satış Adedi: 2

En Yüksek Satış Adedine Sahip Marka: Şafak Ambalaj

Toplam Satış Adedi: 2250

3) I ranked the brands based on their associated colors according to the revenue generated, aiming to identify the brand-color combination with the highest revenue. This analysis serves to highlight the most lucrative brand-color pair, offering insights into consumer preferences and guiding marketing strategies to capitalize on successful brand-color associations. Also, I sorted the return rates by brand to evaluate the performance of each brand in terms of product returns. This analysis helps in identifying brands with

Renk ve Marka Bilgileri (Net Ciro Göre Sıralı):

Renk: Sarı, Marka: Ancare, Net Ciro: 55189.0

Renk: Gri, Marka: Outdoor Outlet, Net Ciro: 15170.0

Renk: Sarı, Marka: SANOR, Net Ciro: 14071.0

Renk: Sarı, Marka: Orion Outdoor, Net Ciro: 11001.0

Renk: Sarı, Marka: Şafak Ambalaj, Net Ciro: 8509.0

Renk: Gri, Marka: MiO, Net Ciro: 7403.0

Renk: Sarı, Marka: Trend Stone, Net Ciro: 6805.0

Renk: Gri, Marka: Şafak Ambalaj, Net Ciro: 4088.0

Renk: Gri, Marka: Outdoor Outlet, Net Ciro: 3915.0

Renk: Gri, Marka: Şafak Ambalaj, Net Ciro: 3790.0

Renk: Sarı, Marka: Webdesepet, Net Ciro: 2497.0

Renk: Sarı, Marka: Ancare, Net Ciro: 2397.0

Renk: Gri, Marka: MiO, Net Ciro: 1562.0

Renk: Gri, Marka: MiO, Net Ciro: 1533.0

Renk: Gri, Marka: Vox Art, Net Ciro: 281.0

Renk: Gri, Marka: Şafak Ambalaj, Net Ciro: 200.53

En Yüksek Net Ciroya Sahip Ürün:
Marka: Ancare, Renk: Sarı, Net Ciro: 55189.0

Markalar ve Ürün İsimleri İptal Oranına Göre Sıralı:

Marka: MiO, İptal Oranı: 14.38

Marka: MiO, İptal Oranı: 4.55

Marka: Ancare, İptal Oranı: 3.6

Marka: Şafak Ambalaj, İptal Oranı: 2.98

Marka: Ancare, İptal Oranı: 2.35

Marka: Şafak Ambalaj, İptal Oranı: 2.0

Marka: Outdoor Outlet, İptal Oranı: 1.96

Marka: MiO, İptal Oranı: 1.2

Marka: SANOR, İptal Oranı: 0.82

Marka: Orion Outdoor, İptal Oranı: 0.72

Marka: Şafak Ambalaj, İptal Oranı: 0.0

Marka: Trend Stone, İptal Oranı: 0.0

Marka: Outdoor Outlet, İptal Oranı: 0.0

Marka: Şafak Ambalaj, İptal Oranı: 0.0

Marka: Webdesepet, İptal Oranı: 0.0

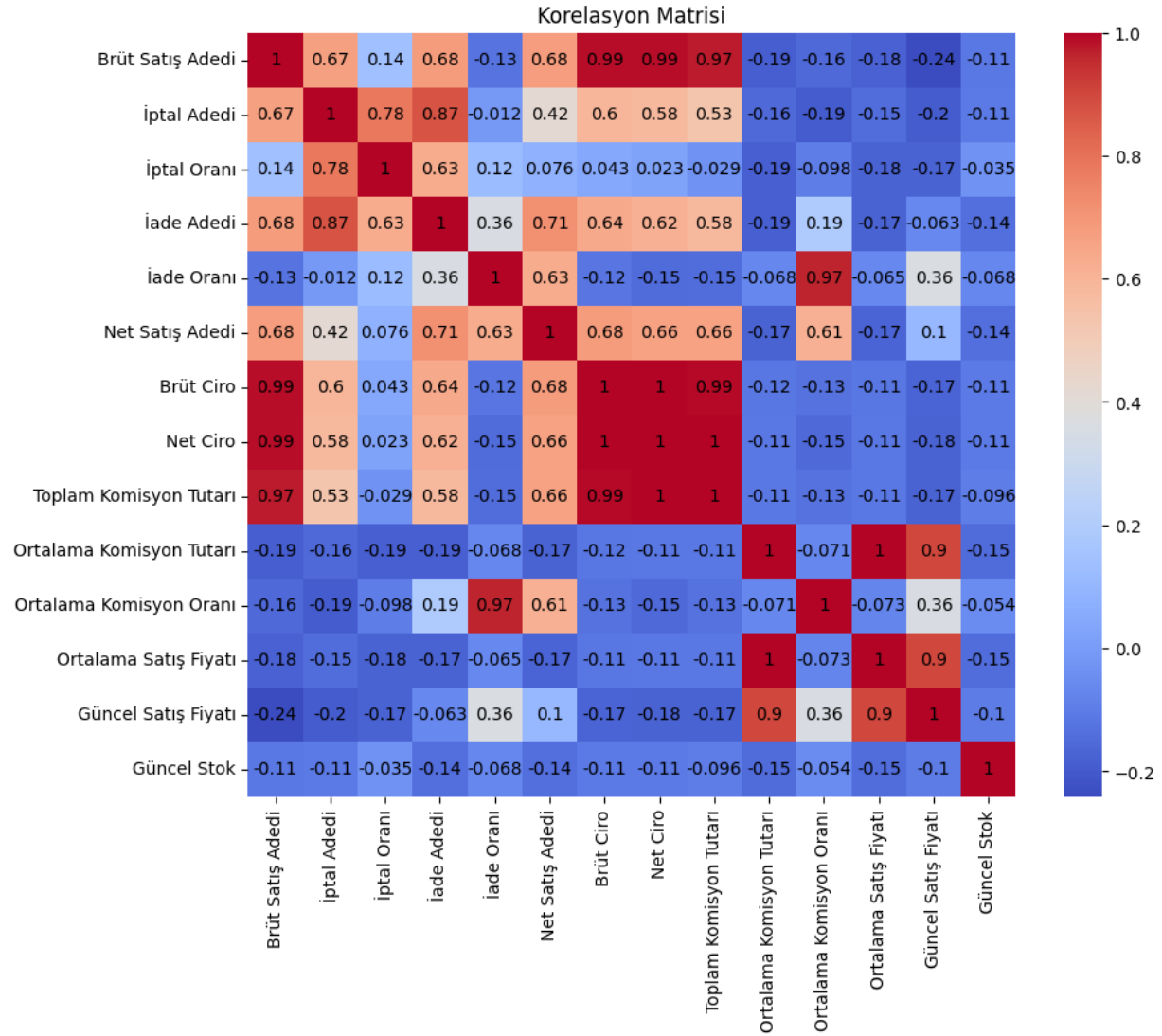
Marka: Vox Art, İptal Oranı: 0.0

En Yüksek İptal Oranına Sahip Ürün:
Marka: MiO, Ürün: Termal Battaniye, İptal Oranı: 14.38

higher or lower return rates, providing insights into potential quality issues, customer satisfaction levels, and areas for improvement in the product lineup or customer service strategies.

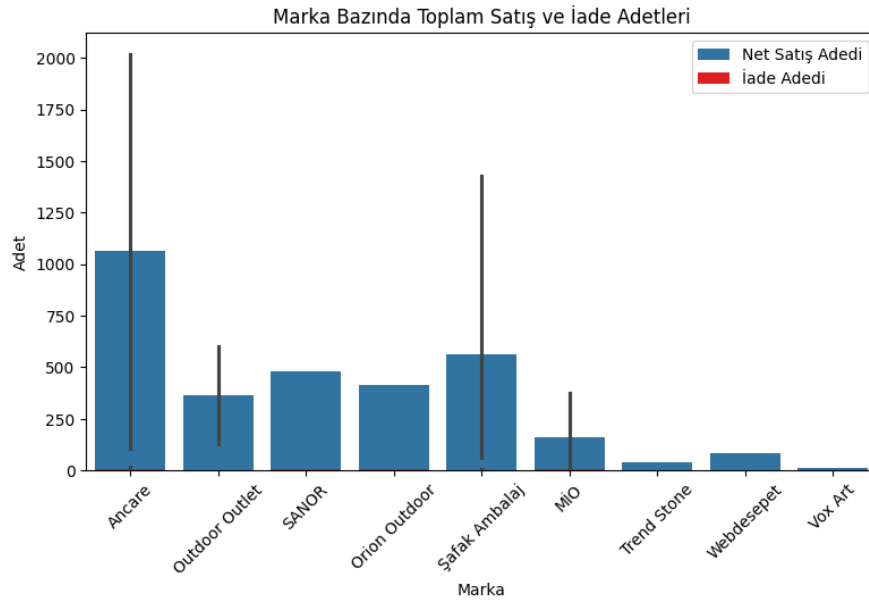
Data Visualization

1-) I created a correlation matrix for all the columns to explore the relationships and dependencies between different metrics such as sales quantity, revenue, return rates, and commission amounts.

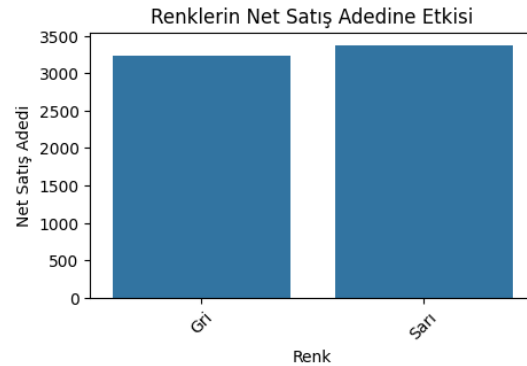
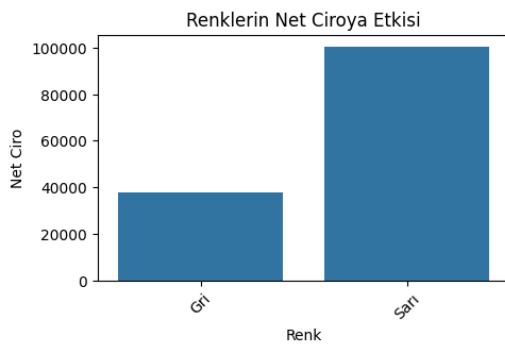


This analysis aims to uncover patterns and associations within the data, providing valuable insights into how various factors interrelate and influence each other. Also, it will be useful to determining hypothesis tests.

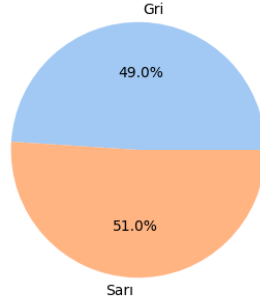
2-) I created a graph depicting sales quantity and return quantity based on brands to visually analyze the performance of each brand in terms of both sales and returns.



3-) I generated three graphs illustrating net sales quantity and net revenue based on colors to visually examine the performance of different colors in the product lineup. This graphical representation allows for a quick assessment of how each color contributes to both sales quantity and revenue.

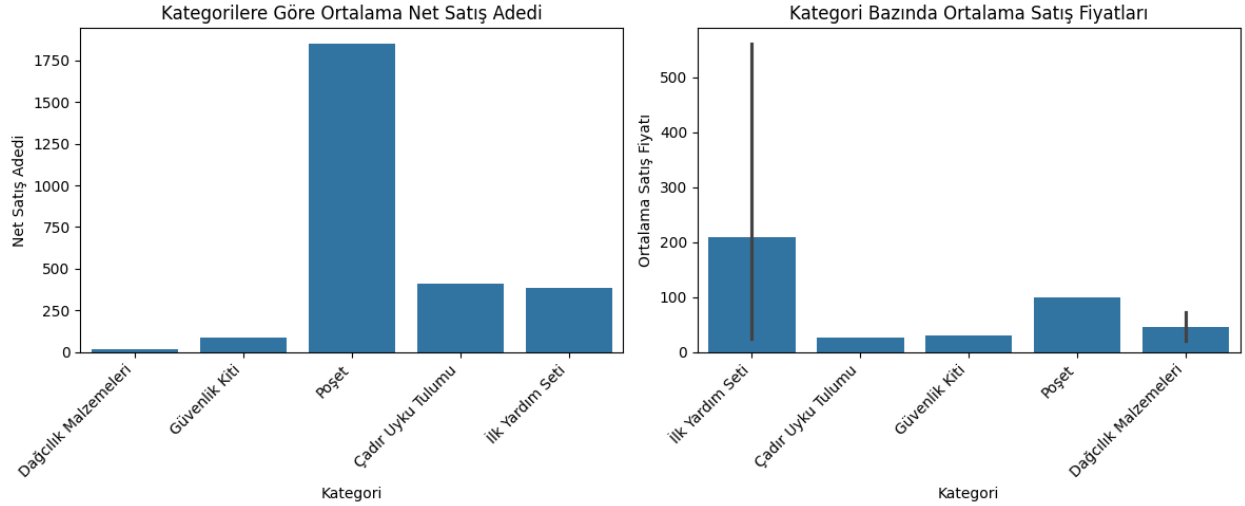


Renklerin Satış Adedine Göre Dağılımı



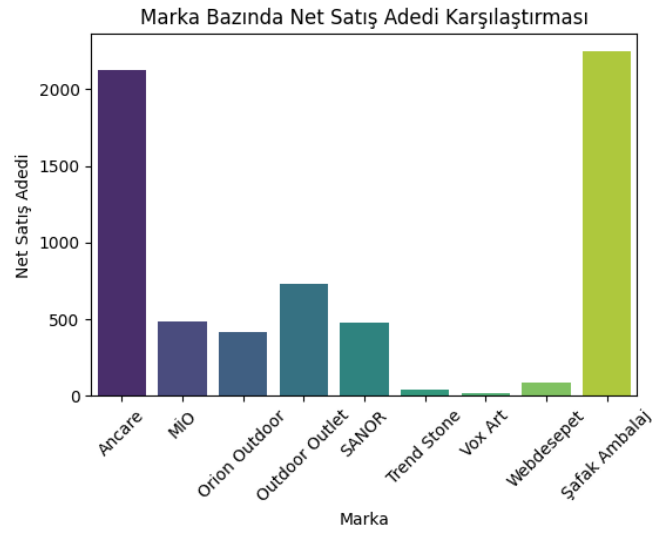
The objective is to identify color-related trends and patterns that may influence customer preferences and purchasing behavior.

4-) I created a graph depicting net sales quantity and net revenue based on product categories to visually analyze the performance of each category in terms of sales and revenue. This graphical representation enables a quick assessment of how different categories contribute to overall sales and revenue.

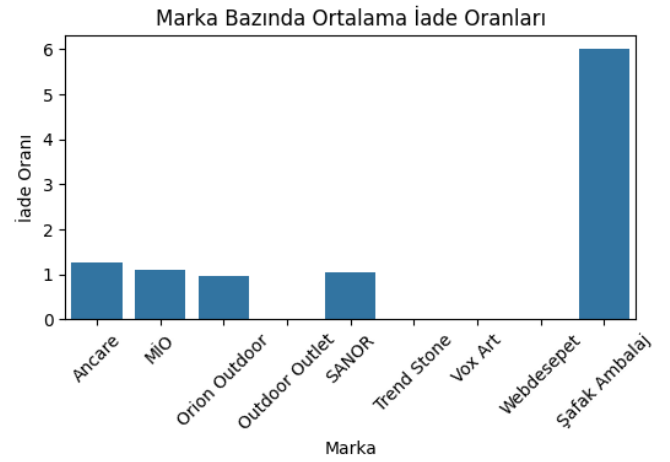
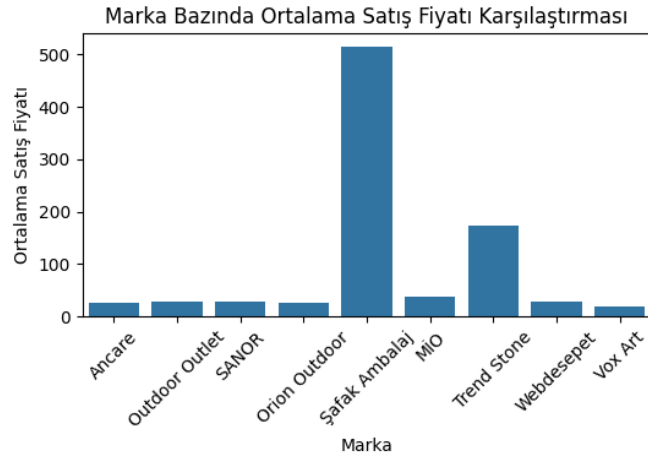


The goal is to identify category-specific trends and patterns, aiding in strategic decision-making related to product development, marketing, and inventory management.

5-) I generated a graph showcasing net sales quantity and net revenue based on brands to visually assess the performance of each brand in terms of both sales volume and revenue.



This graphical representation enables a quick assessment of how different categories contribute to overall sales and revenue. Then I printed the graphs that showing returns rate and average sales price according to brands.



Hypothesis Tests

I started to do hypothesis test, I generate a code that find the p value and if p value is greater 0.05 hypothesis will be accepted otherwise it will be rejected.

```
1 from scipy.stats import pearsonr
2
3 # H0: İki değişken arasında korelasyon yoktur.
4 # H1: İki değişken arasında korelasyon vardır.
5
6 # Örnek: İki değişken seçilmiş olsun, 'Komisyon_Tutari' ve 'Guncel_Satis_Fiyati'
7 komisyon_tutari = data['Toplam Komisyon Tutarı'].dropna()
8 guncel_satis_fiyati = data['Güncel Satış Fiyatı'].dropna()
9
10 # Pearson Korelasyon Testi uygula
11 corr_coefficient, p_value = pearsonr(komisyon_tutari, guncel_satis_fiyati)
12
13 print(f"Pearson Korelasyon Katsayısı: {corr_coefficient}")
14 print(f"P-değeri: {p_value}")
15
16 if p_value < 0.05:
17     print("H0 reddedildi. Komisyon Tutarı ile Güncel Satış Fiyatı arasında istatistiksel olarak anlamlı bir korelasyon vardır.")
18 else:
19     print("H0 kabul edildi. Komisyon Tutarı ile Güncel Satış Fiyatı arasında istatistiksel olarak anlamlı bir korelasyon yoktur.")
20
```

Pearson Korelasyon Katsayısı: -0.17335633878355025
P-değeri: 0.5208300795263241
H0 kabul edildi. Komisyon Tutarı ile Güncel Satış Fiyatı arasında istatistiksel olarak anlamlı bir korelasyon yoktur.

Since, p value is 0.52 which is greater than 0.05. H0 is accepted, there is no correlation between commission rate and sale price.

```
1 from scipy.stats import pearsonr
2
3 # H0: İki değişken arasında korelasyon yoktur.
4 # H1: İki değişken arasında korelasyon vardır.
5
6 # Örnek: İki değişken seçilmiş olsun, 'Net_Ciro' ve 'Toplam_Komisyon_Tutari'
7 net_ciro = data['Net Ciro'].dropna()
8 toplam_komisyon_tutari = data['Toplam Komisyon Tutarı'].dropna()
9
10 # Pearson Korelasyon Testi uygula
11 corr_coefficient, p_value = pearsonr(net_ciro, toplam_komisyon_tutari)
12
13 print(f"Pearson Korelasyon Katsayısı: {corr_coefficient}")
14 print(f"P-değeri: {p_value}")
15
16 if p_value < 0.05:
17     print("H0 reddedildi. Net Ciro ile Toplam Komisyon Tutarı arasında istatistiksel olarak anlamlı bir korelasyon vardır.")
18 else:
19     print("H0 kabul edildi. Net Ciro ile Toplam Komisyon Tutarı arasında istatistiksel olarak anlamlı bir korelasyon yoktur.")
20
```

Pearson Korelasyon Katsayısı: 0.9952363739497616
P-değeri: 1.4738900314096715e-15
H0 reddedildi. Net Ciro ile Toplam Komisyon Tutarı arasında istatistiksel olarak anlamlı bir korelasyon vardır.

P value is too lower than 0.05 this means that H0 will be rejected. Thus there is a high correlation between net revenue and total commission.

```

1 from scipy.stats import ttest_ind
2
3 # H0: İptal Adedi ile Net Ciro arasında istatistiksel olarak anlamlı bir fark yoktur.
4 # H1: İptal Adedi ile Net Ciro arasında istatistiksel olarak anlamlı bir fark vardır.
5
6 # Örnek: İki örneklem seçilmiş olsun, 'İptal_Adedi' ve 'Net_Ciro'
7 iptal_adedi = data['İptal Adedi'].sample(frac=1, random_state=42)
8 net_ciro = data['Net Ciro'].sample(frac=1, random_state=42)
9
10 # İki Örneklem T-Testi uygula
11 stat, p_value = ttest_ind(iptal_adedi, net_ciro, equal_var=False)
12
13 print(f"T-Test İstatistiği: {stat}")
14 print(f"P-değeri: {p_value}")
15
16 if p_value < 0.05:
17     print("H0 reddedildi. İptal Adedi ile Net Ciro arasında istatistiksel olarak anlamlı bir fark vardır.")
18 else:
19     print("H0 kabul edildi. İptal Adedi ile Net Ciro arasında istatistiksel olarak anlamlı bir fark yoktur.")
20
T-Test İstatistiği: -2.6083138805659782
P-değeri: 0.019768943218710273
H0 reddedildi. İptal Adedi ile Net Ciro arasında istatistiksel olarak anlamlı bir fark vardır.

```

P value of this hypothesis test is 0.019 which is lower than 0.05. H0 is rejected there is a difference between rejection number and net revenue.

```

1 from scipy.stats import ttest_ind
2
3 # H0: Ortalama satış fiyatı ile iade oranı arasında istatistiksel olarak anlamlı bir fark yoktur.
4 # H1: Ortalama satış fiyatı ile iade oranı arasında istatistiksel olarak anlamlı bir fark vardır.
5
6 # Örnek: İki örneklem seçilmiş olsun, 'Ortalama_Satis_Fiyati' ve 'Iade_Orani'
7 ortalama_satis_fiyati = data['Ortalama Satış Fiyatı'].sample(frac=1, random_state=42)
8 iade_orani = data['İade Oranı'].sample(frac=1, random_state=42)
9
10 # İki Örneklem T-Testi uygula
11 stat, p_value = ttest_ind(ortalama_satis_fiyati, iade_orani, equal_var=False)
12
13 print(f"T-Test İstatistiği: {stat}")
14 print(f"P-değeri: {p_value}")
15
16 if p_value < 0.05:
17     print("H0 reddedildi. Ortalama satış fiyatı ile iade oranı arasında istatistiksel olarak anlamlı bir fark vardır.")
18 else:
19     print("H0 kabul edildi. Ortalama satış fiyatı ile iade oranı arasında istatistiksel olarak anlamlı bir fark yoktur.")
20
T-Test İstatistiği: 1.3571054634094462
P-değeri: 0.19481285999018721
H0 kabul edildi. Ortalama satış fiyatı ile iade oranı arasında istatistiksel olarak anlamlı bir fark yoktur.

```

P value of this test 0.19 which is greater than 0.05. H0 is accepted, there is no difference between avg sale price and return rate.

I selected the most cancelled brands among others and I used them to test whether there is statistically significant difference between cancellation rates by brand or not.


```

1 from scipy.stats import f_oneway
2
3 # H0: Marka bazında iptal oranları arasında istatistiksel olarak anlamlı bir fark yoktur.
4 # H1: Marka bazında iptal oranları arasında istatistiksel olarak anlamlı bir fark vardır.
5
6 # Örnek: İki marka seçilmiş olsun, 'Marka_A' ve 'Marka_B'
7 marka_A iptal_orani = data[data['Marka'] == 'Ancare']['İptal Oranı'].sample(frac=1, random_state=42)
8 marka_B iptal_orani = data[data['Marka'] == 'Şafak Ambalaj']['İptal Oranı'].sample(frac=1, random_state=42)
9
10 # ANOVA (One-way Analysis of Variance) testi uygula
11 stat, p_value = f_oneway(marka_A iptal_orani, marka_B iptal_orani)
12
13 print(f"ANOVA İstatistiği: {stat}")
14 print(f"P-değeri: {p_value}")
15
16 if p_value < 0.05:
17     print("H0 reddedildi. Marka bazında iptal oranları arasında istatistiksel olarak anlamlı bir fark vardır.")
18 else:
19     print("H0 kabul edildi. Marka bazında iptal oranları arasında istatistiksel olarak anlamlı bir fark yoktur.")
20
ANOVA İstatistiği: 2.139251674696723
P-değeri: 0.21739564227631974
H0 kabul edildi. Marka bazında iptal oranları arasında istatistiksel olarak anlamlı bir fark yoktur.

```

So, p value is 0.21 then accept the H0. There is no statistically significant difference between cancellation rates by brand.

Regression

I build a regression model for net sales amount, I determined the X values according to their influence on the sales amount which are net revenue and average sales price.

```
1 import pandas as pd
2 from sklearn.model_selection import train_test_split
3 from sklearn.linear_model import LinearRegression
4 from sklearn.metrics import mean_squared_error
5
6
7 # Eksik değerleri temizle
8 data = data.dropna()
9
10 # Bağımsız değişkenler (X) ve bağımlı değişken (y) seçimi
11 X = data[['Net Ciro', 'Ortalama Satış Fiyatı']]
12 y = data['Net Satış Adedi']
13
14 # Veriyi eğitim ve test setlerine bölelim
15 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
16
17 # Model oluşturup eğitelim
18 model = LinearRegression()
19 model.fit(X_train, y_train)
20
21 # Test seti üzerinde tahmin yapalım
22 y_pred = model.predict(X_test)
23
24 # Hata ölçümü yapalım (örneğin, ortalama karesel hata)
25 mse = mean_squared_error(y_test, y_pred)
26 print(f'Mean Squared Error: {mse}')
27
```

Mean Squared Error: 1354082.0482331477

Then I try to predict the incoming future demand, from the regression model.

```
15 # Bağımsız değişkenleri (X) ve bağımlı değişkeni (y) seç
16 X = data[['Brüt Ciro', 'Ortalama Satış Fiyatı', 'Marka']]
17 y = data['Net Satış Adedi']
18
19 # Eğitim ve test veri setlerini oluştur
20 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
21
22 # Marka sütununu one-hot encoding yap
23 preprocessor = ColumnTransformer(
24     transformers=[('cat', OneHotEncoder(), ['Marka'])],
25     remainder='passthrough'
26 )
27
28 # Modeli oluştur ve eğit
29 model = Pipeline(steps=[('preprocessor', preprocessor),
30     ('regressor', LinearRegression())])
31
32 model.fit(X_train, y_train)
33
34 # Test veri seti üzerinde tahmin yap
35 y_pred = model.predict(X_test)
36
37 # Model performansını değerlendir
38 mse = mean_squared_error(y_test, y_pred)
39 print(f'Mean Squared Error: {mse}')
40
41 # Tahminleri görselleştir
42 plt.scatter(y_test, y_pred)
43 plt.xlabel('Gerçek Değerler')
44 plt.ylabel('Tahmin Edilen Değerler')
45 plt.title('Net Satış Adedi Tahminleri (Marka Bazında)')
46 plt.show()
47
```

