

Problem

Customer Segmentation Analysis

Gezinomi wants to create new level-based sales definitions using some features of the sales it makes and create segments according to these new sales **BUSINESS PROBLEM, DATA SET, VARIABLES, TASKS** definitions and estimate how much new customers can earn on average for the company according to these segments.

For example: It is desired to determine how much a customer who wants to go to an All-Inclusive hotel from Antalya during a busy period can earn on average.

```
In [42]: import pandas as pd
pd.set_option('display.max_columns', None)
pd.set_option('display.float_format', lambda x: '%.2f' % x)
```

```
In [2]: df = pd.read_excel('gezinomi.xlsx')
```

```
In [3]: df.head()
```

```
Out[3]:
```

	SaleId	SaleDate	CheckInDate	Price	ConceptName	SaleCityName	CInDay	SaleCheckInD.
0	415122	2022-12-03	2022-12-03	79.30	Herşey Dahil	Antalya	Saturday	
1	415103	2022-12-03	2022-12-03	45.97	Yarım Pansiyon	Antalya	Saturday	
2	404034	2022-09-12	2022-09-13	77.84	Herşey Dahil	Antalya	Tuesday	
3	415094	2022-12-03	2022-12-10	222.71	Yarım Pansiyon	İzmir	Saturday	
4	414951	2022-12-01	2022-12-03	140.48	Yarım Pansiyon	İzmir	Saturday	

```
In [4]: df.shape
```

```
Out[4]: (59164, 9)
```

In [5]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 59164 entries, 0 to 59163
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   SaleId                59164 non-null  int64
1   SaleDate              59164 non-null  datetime64[ns]
2   CheckInDate           59164 non-null  datetime64[ns]
3   Price                 59151 non-null  float64
4   ConceptName           59164 non-null  object
5   SaleCityName          59164 non-null  object
6   CInDay               59164 non-null  object
7   SaleCheckInDayDiff    59164 non-null  int64
8   Seasons               59164 non-null  object
dtypes: datetime64[ns](2), float64(1), int64(2), object(4)
memory usage: 4.1+ MB
```

In [6]: df['SaleCityName'].nunique()

Out[6]: 6

In [7]: df['SaleCityName'].value_counts()

Out[7]: SaleCityName

Antalya	31649
Muğla	10662
Aydın	10646
Diğer	3245
İzmir	2507
Girne	455

Name: count, dtype: int64

In [8]: df['ConceptName'].nunique()

Out[8]: 3

In [9]: df['ConceptName'].value_counts()

Out[9]: ConceptName

Herşey Dahil	53186
Yarım Pansiyon	3559
Oda + Kahvaltı	2419

Name: count, dtype: int64

```
In [10]: df.groupby('SaleCityName').agg({'Price': 'sum'})
```

```
Out[10]:
```

	Price
SaleCityName	
Antalya	2041911.10
Aydın	573296.01
Diğer	154572.29
Girne	27065.03
Muğla	665842.21
İzmir	165934.83

```
In [11]: df.groupby('ConceptName').agg({'Price': 'sum'})
```

```
Out[11]:
```

	Price
ConceptName	
Herşey Dahil	3332910.77
Oda + Kahvaltı	121308.35
Yarım Pansiyon	174402.35

```
In [12]: df.groupby('SaleCityName').agg({'Price': 'mean'})
```

```
Out[12]:
```

	Price
SaleCityName	
Antalya	64.52
Aydın	53.86
Diğer	47.71
Girne	59.48
Muğla	62.46
İzmir	66.27

```
In [13]: df.groupby('ConceptName').agg({'Price': 'mean'})
```

```
Out[13]:
```

	Price
ConceptName	
Herşey Dahil	62.67
Oda + Kahvaltı	50.25
Yarım Pansiyon	49.03

```
In [16]: df.groupby(['SaleCityName', 'ConceptName']).agg({'Price': 'mean'})
```

```
Out[16]:
```

		Price
SaleCityName	ConceptName	
	Herşey Dahil	64.52
Antalya	Oda + Kahvaltı	63.50
	Yarım Pansiyon	67.19
	Herşey Dahil	54.00
Aydın	Oda + Kahvaltı	34.46
	Yarım Pansiyon	30.02
	Herşey Dahil	84.77
Diğer	Oda + Kahvaltı	37.60
	Yarım Pansiyon	42.11
	Herşey Dahil	97.68
Girne	Oda + Kahvaltı	39.78
	Yarım Pansiyon	53.25
	Herşey Dahil	63.02
Muğla	Oda + Kahvaltı	59.04
	Yarım Pansiyon	45.12
	Herşey Dahil	74.70
İzmir	Oda + Kahvaltı	41.32
	Yarım Pansiyon	59.61

```
In [20]: bins = [-1, 7, 30, 90, df['SaleCheckInDayDiff'].max()]
labels = ['Last Minutes', 'Potential Planners', 'Planners', 'Early Bookers']
```

```
In [21]: df['EB_Score'] = pd.cut(df['SaleCheckInDayDiff'], bins, labels = labels)
```

```
In [22]: df.head(50).to_excel('eb_score.xlsx', index = False)
```

```
In [25]: df.groupby(['SaleCityName', 'ConceptName', 'EB_Score']).agg({'Price': ['mean',
```

```
Out[25]:
```

			Price	
			mean	count
SaleCityName	ConceptName	EB_Score		
Antalya	Herşey Dahil	Last Minutes	62.75	14148
		Potential Planners	64.90	8874
		Planners	67.88	4490
		Early Bookers	66.49	3281
	Oda + Kahvaltı	Last Minutes	65.35	503
...
İzmir	Yarım Pansiyon	Early Bookers	66.93	3
		Last Minutes	57.79	636
		Potential Planners	64.80	198
		Planners	60.51	47
		Early Bookers	65.91	14

72 rows × 2 columns

```
In [27]: df.groupby(['SaleCityName', 'ConceptName', 'Seasons']).agg({'Price': ['mean',
```

Out[27]:

SaleCityName	ConceptName	Seasons	Price	
			mean	count
Antalya	Herşey Dahil	High	64.92	27126
		Low	61.55	3667
	Oda + Kahvaltı	High	66.27	303
		Low	60.67	295
	Yarım Pansiyon	High	73.26	118
		Low	62.00	138
	Herşey Dahil	High	54.95	10103
		Low	33.68	473
Aydın	Oda + Kahvaltı	High	30.39	27
		Low	44.45	11
	Yarım Pansiyon	High	32.63	20
		Low	25.27	11
	Herşey Dahil	High	83.79	364
		Low	87.31	141
	Oda + Kahvaltı	High	34.84	349
		Low	39.96	408
Diğer	Yarım Pansiyon	High	39.73	516
		Low	42.95	1462
	Herşey Dahil	High	103.94	55
		Low	90.94	51
	Oda + Kahvaltı	High	39.48	66
		Low	40.04	73
	Yarım Pansiyon	High	58.38	100
		Low	48.58	110
Girne	Herşey Dahil	High	63.38	9582
		Low	48.30	237
	Oda + Kahvaltı	High	61.26	592
		Low	37.77	62
	Yarım Pansiyon	High	45.82	177
		Low	32.68	10

SaleCityName	ConceptName	Seasons	Price	
			mean	count
İzmir	Herşey Dahil	High	74.75	1234
		Low	74.31	147
	Oda + Kahvaltı	High	54.14	86
		Low	33.56	142
	Yarım Pansiyon	High	87.66	118
		Low	55.35	777

In [28]: `df.groupby(['SaleCityName', 'ConceptName', 'CInDay']).agg({'Price': ['mean', 'count']})`

Out[28]:

SaleCityName	ConceptName	CInDay	Price	
			mean	count
Antalya	Herşey Dahil	Friday	62.66	4136
		Monday	63.26	6831
		Saturday	64.42	4741
		Sunday	65.85	3818
		Thursday	62.89	3898
...
İzmir	Yarım Pansiyon	Saturday	52.50	315
		Sunday	98.93	68
		Thursday	52.48	91
		Tuesday	56.14	73
		Wednesday	54.06	86

126 rows × 5 columns

In [29]: `agg_df = df.groupby(['SaleCityName', 'ConceptName', 'Seasons']).agg({'Price': ['mean', 'count']})`


```
In [31]: agg_df.head(15)
```

```
Out[31]:
```

	SaleCityName	ConceptName	Seasons	Price
	Girne	Herşey Dahil	High	103.94
			Low	90.94
	İzmir	Yarım Pansiyon	High	87.66
			Low	87.31
	Diğer	Herşey Dahil	High	83.79
			High	74.75
	İzmir	Herşey Dahil	Low	74.31
		Yarım Pansiyon	High	73.26
	Antalya	Oda + Kahvaltı	High	66.27
		Herşey Dahil	High	64.92
	Muğla	Herşey Dahil	High	63.38
		Yarım Pansiyon	Low	62.00
	Antalya	Herşey Dahil	Low	61.55
	Muğla	Oda + Kahvaltı	High	61.26
	Antalya	Oda + Kahvaltı	Low	60.67

```
In [32]: agg_df.reset_index(inplace = True)
```

```
In [33]: agg_df.head()
```

```
Out[33]:
```

	SaleCityName	ConceptName	Seasons	Price
0	Girne	Herşey Dahil	High	103.94
1	Girne	Herşey Dahil	Low	90.94
2	İzmir	Yarım Pansiyon	High	87.66
3	Diğer	Herşey Dahil	Low	87.31
4	Diğer	Herşey Dahil	High	83.79

```
In [35]: agg_df['sales_level_based'] = agg_df[['SaleCityName', 'ConceptName', 'Seasons']
```

```
In [37]: agg_df['Segment'] = pd.qcut(agg_df['Price'], 4, labels=['D', 'C', 'B', 'A'])
agg_df.head(25)
```

```
Out[37]:
```

	SaleCityName	ConceptName	Seasons	Price	sales_level_based	Segment
0	Girne	Herşey Dahil	High	103.94	GIRNE_HERŞEY DAHIL_HIGH	A
1	Girne	Herşey Dahil	Low	90.94	GIRNE_HERŞEY DAHIL_LOW	A
2	İzmir	Yarım Pansiyon	High	87.66	İZMIR_YARIM PANSİYON_HIGH	A
3	Diğer	Herşey Dahil	Low	87.31	DIĞER_HERŞEY DAHIL_LOW	A
4	Diğer	Herşey Dahil	High	83.79	DIĞER_HERŞEY DAHIL_HIGH	A
5	İzmir	Herşey Dahil	High	74.75	İZMIR_HERŞEY DAHIL_HIGH	A
6	İzmir	Herşey Dahil	Low	74.31	İZMIR_HERŞEY DAHIL_LOW	A
7	Antalya	Yarım Pansiyon	High	73.26	ANTALYA_YARIM PANSİYON_HIGH	A
8	Antalya	Oda + Kahvaltı	High	66.27	ANTALYA_ODA + KAHVALTI_HIGH	A
9	Antalya	Herşey Dahil	High	64.92	ANTALYA_HERŞEY DAHIL_HIGH	B
10	Muğla	Herşey Dahil	High	63.38	MUĞLA_HERŞEY DAHIL_HIGH	B
11	Antalya	Yarım Pansiyon	Low	62.00	ANTALYA_YARIM PANSİYON_LOW	B
12	Antalya	Herşey Dahil	Low	61.55	ANTALYA_HERŞEY DAHIL_LOW	B
13	Muğla	Oda + Kahvaltı	High	61.26	MUĞLA_ODA + KAHVALTI_HIGH	B
14	Antalya	Oda + Kahvaltı	Low	60.67	ANTALYA_ODA + KAHVALTI_LOW	B
15	Girne	Yarım Pansiyon	High	58.38	GIRNE_YARIM PANSİYON_HIGH	B
16	İzmir	Yarım Pansiyon	Low	55.35	İZMIR_YARIM PANSİYON_LOW	B
17	Aydın	Herşey Dahil	High	54.95	AYDIN_HERŞEY DAHIL_HIGH	B
18	İzmir	Oda + Kahvaltı	High	54.14	İZMIR_ODA + KAHVALTI_HIGH	C
19	Girne	Yarım Pansiyon	Low	48.58	GIRNE_YARIM PANSİYON_LOW	C
20	Muğla	Herşey Dahil	Low	48.30	MUĞLA_HERŞEY DAHIL_LOW	C
21	Muğla	Yarım Pansiyon	High	45.82	MUĞLA_YARIM PANSİYON_HIGH	C
22	Aydın	Oda + Kahvaltı	Low	44.45	AYDIN_ODA + KAHVALTI_LOW	C
23	Diğer	Yarım Pansiyon	Low	42.95	DIĞER_YARIM PANSİYON_LOW	C
24	Girne	Oda + Kahvaltı	Low	40.04	GIRNE_ODA + KAHVALTI_LOW	C

```
In [38]: agg_df.groupby('Segment').agg({'Price': ['mean', 'max', 'sum']})
```

Out[38]:

	Price		
	mean	max	sum
Segment			
D	33.37	39.48	300.30
C	44.89	54.14	403.99
B	60.27	64.92	542.47
A	82.47	103.94	742.21

```
In [39]: agg_df.sort_values(by='Price')
```

Out[39]:

	SaleCityName	ConceptName	Seasons	Price	sales_level_based	Segment
35	Aydın	Yarım Pansiyon	Low	25.27	AYDIN_YARIM PANSIYON_LOW	D
34	Aydın	Oda + Kahvaltı	High	30.39	AYDIN_ODA + KAHVALTI_HIGH	D
33	Aydın	Yarım Pansiyon	High	32.63	AYDIN_YARIM PANSIYON_HIGH	D
32	Muğla	Yarım Pansiyon	Low	32.68	MUĞLA_YARIM PANSIYON_LOW	D
31	İzmir	Oda + Kahvaltı	Low	33.56	İZMİR_ODA + KAHVALTI_LOW	D
30	Aydın	Herşey Dahil	Low	33.68	AYDIN_HERŞEY DAHIL_LOW	D
29	Diğer	Oda + Kahvaltı	High	34.84	DIĞER_ODA + KAHVALTI_HIGH	D
28	Muğla	Oda + Kahvaltı	Low	37.77	MUĞLA_ODA + KAHVALTI_LOW	D
27	Girne	Oda + Kahvaltı	High	39.48	GIRNE_ODA + KAHVALTI_HIGH	D
26	Diğer	Yarım Pansiyon	High	39.73	DIĞER_YARIM PANSIYON_HIGH	C
25	Diğer	Oda + Kahvaltı	Low	39.96	DIĞER_ODA + KAHVALTI_LOW	C
24	Girne	Oda + Kahvaltı	Low	40.04	GIRNE_ODA + KAHVALTI_LOW	C
23	Diğer	Yarım Pansiyon	Low	42.95	DIĞER_YARIM PANSIYON_LOW	C
22	Aydın	Oda + Kahvaltı	Low	44.45	AYDIN_ODA + KAHVALTI_LOW	C
21	Muğla	Yarım Pansiyon	High	45.82	MUĞLA_YARIM PANSIYON_HIGH	C
20	Muğla	Herşey Dahil	Low	48.30	MUĞLA_HERŞEY DAHIL_LOW	C
19	Girne	Yarım Pansiyon	Low	48.58	GIRNE_YARIM PANSIYON_LOW	C
18	İzmir	Oda + Kahvaltı	High	54.14	İZMİR_ODA + KAHVALTI_HIGH	C
17	Aydın	Herşey Dahil	High	54.95	AYDIN_HERŞEY DAHIL_HIGH	B
16	İzmir	Yarım Pansiyon	Low	55.35	İZMİR_YARIM PANSIYON_LOW	B
15	Girne	Yarım Pansiyon	High	58.38	GIRNE_YARIM PANSIYON_HIGH	B
14	Antalya	Oda + Kahvaltı	Low	60.67	ANTALYA_ODA + KAHVALTI_LOW	B
13	Muğla	Oda + Kahvaltı	High	61.26	MUĞLA_ODA + KAHVALTI_HIGH	B
12	Antalya	Herşey Dahil	Low	61.55	ANTALYA_HERŞEY DAHIL_LOW	B
11	Antalya	Yarım Pansiyon	Low	62.00	ANTALYA_YARIM PANSIYON_LOW	B
10	Muğla	Herşey Dahil	High	63.38	MUĞLA_HERŞEY DAHIL_HIGH	B
9	Antalya	Herşey Dahil	High	64.92	ANTALYA_HERŞEY DAHIL_HIGH	B
8	Antalya	Oda + Kahvaltı	High	66.27	ANTALYA_ODA + KAHVALTI_HIGH	A
7	Antalya	Yarım Pansiyon	High	73.26	ANTALYA_YARIM PANSIYON_HIGH	A
6	İzmir	Herşey Dahil	Low	74.31	İZMİR_HERŞEY DAHIL_LOW	A

	SaleCityName	ConceptName	Seasons	Price	sales_level_based	Segment
5	İzmir	Herşey Dahil	High	74.75	İZMİR_HERŞEY DAHIL_HIGH	A
4	Diğer	Herşey Dahil	High	83.79	DİĞER_HERŞEY DAHIL_HIGH	A
3	Diğer	Herşey Dahil	Low	87.31	DİĞER_HERŞEY DAHIL_LOW	A
2	İzmir	Yarım Pansiyon	High	87.66	İZMİR_YARIM PANSİYON_HIGH	A
1	Girne	Herşey Dahil	Low	90.94	GİRNE_HERŞEY DAHIL_LOW	A
0	Girne	Herşey Dahil	High	103.94	GİRNE_HERŞEY DAHIL_HIGH	A

```
In [41]: new_user = "ANTALYA_HERŞEY DAHIL_HIGH"
agg_df[agg_df['sales_level_based'] == new_user]
```

```
Out[41]:
```

	SaleCityName	ConceptName	Seasons	Price	sales_level_based	Segment
9	Antalya	Herşey Dahil	High	64.92	ANTALYA_HERŞEY DAHIL_HIGH	B

```
In [ ]:
```