

Exercise 3.2

1. Load the data into a data frame.

```
In [13]: import pandas as pd
df_dodgers=pd.read_csv('dodgers-2022.csv')
df_dodgers.head()
```

```
Out[13]:
```

	month	day	attend	day_of_week	opponent	temp	skies	day_night	cap	shirt	fireworks	bobblehead
0	APR	10	56000	Tuesday	Pirates	67	Clear	Day	NO	NO	NO	NO
1	APR	11	29729	Wednesday	Pirates	58	Cloudy	Night	NO	NO	NO	NO
2	APR	12	28328	Thursday	Pirates	57	Cloudy	Night	NO	NO	NO	NO
3	APR	13	31601	Friday	Padres	54	Cloudy	Night	NO	NO	YES	NO
4	APR	14	46549	Saturday	Padres	57	Cloudy	Night	NO	NO	NO	NO

2. Check the average attendance based on the day of the week.

```
In [14]: # Average attendance by week in ascending order
df_dodgers.groupby('day_of_week')['attend'].mean().sort_values()
```

```
Out[14]: day_of_week
Monday      34965.666667
Wednesday   37585.166667
Friday       40116.923077
Thursday     40407.400000
Sunday       42268.846154
Saturday     43072.923077
Tuesday      47741.230769
Name: attend, dtype: float64
```

3. Check the average attendance based on the opponent.

```
In [15]: # Average attendance by opponent in ascending order
df_dodgers.groupby('opponent')['attend'].mean().sort_values()
```

```
Out[15]: opponent
Braves      32245.000000
Brewers     35358.750000
Astros      35383.333333
Pirates     38019.000000
Giants      39296.333333
Snakes      39315.444444
Rockies     39631.222222
Reds        40649.000000
Marlins     40665.333333
Cardinals   40853.285714
Phillies    41897.000000
Padres      42092.222222
Cubs        44206.666667
White Sox   46382.000000
Nationals   49267.333333
Mets        49586.250000
Angels      49777.333333
Name: attend, dtype: float64
```

4. Check the average attendance based on skies.

```
In [16]: # Average attendance by skies in ascending order
df_dodgers.groupby('skies')['attend'].mean().sort_values()

Out[16]: skies
Cloudy    38791.315789
Clear     41729.209677
Name: attend, dtype: float64
```

5. Check the average attendance based on day/night

```
In [17]: # Average attendance by day/night in ascending order
df_dodgers.groupby('day_night')['attend'].mean().sort_values()

Out[17]: day_night
Night    40868.893939
Day      41793.266667
Name: attend, dtype: float64
```

Based on the above stats it seems that the chances of the attendance are high when dodgers play against Angels/Mets/Nationals (as the avg. attendance is high when they play them) on a Tuesday (surprisingly the average attendance is high on Tuesdays) anytime (as there's only a little difference in avg. attendance between day and night) and when the skies are clear. I ignored the fields cap, shirt, bobble head and fireworks as I assume that they may not have any effect on a match.