In [27]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inLine
import warnings
warnings.filterwarnings('ignore')
#import statsmodels.api as sm
```

In [28]:
```
#Data load
dat = pd.read_csv("dodgers-2022.csv")
print(dat.shape)
dat.head(5)
```

```
(81, 12)
```

Out[28]:

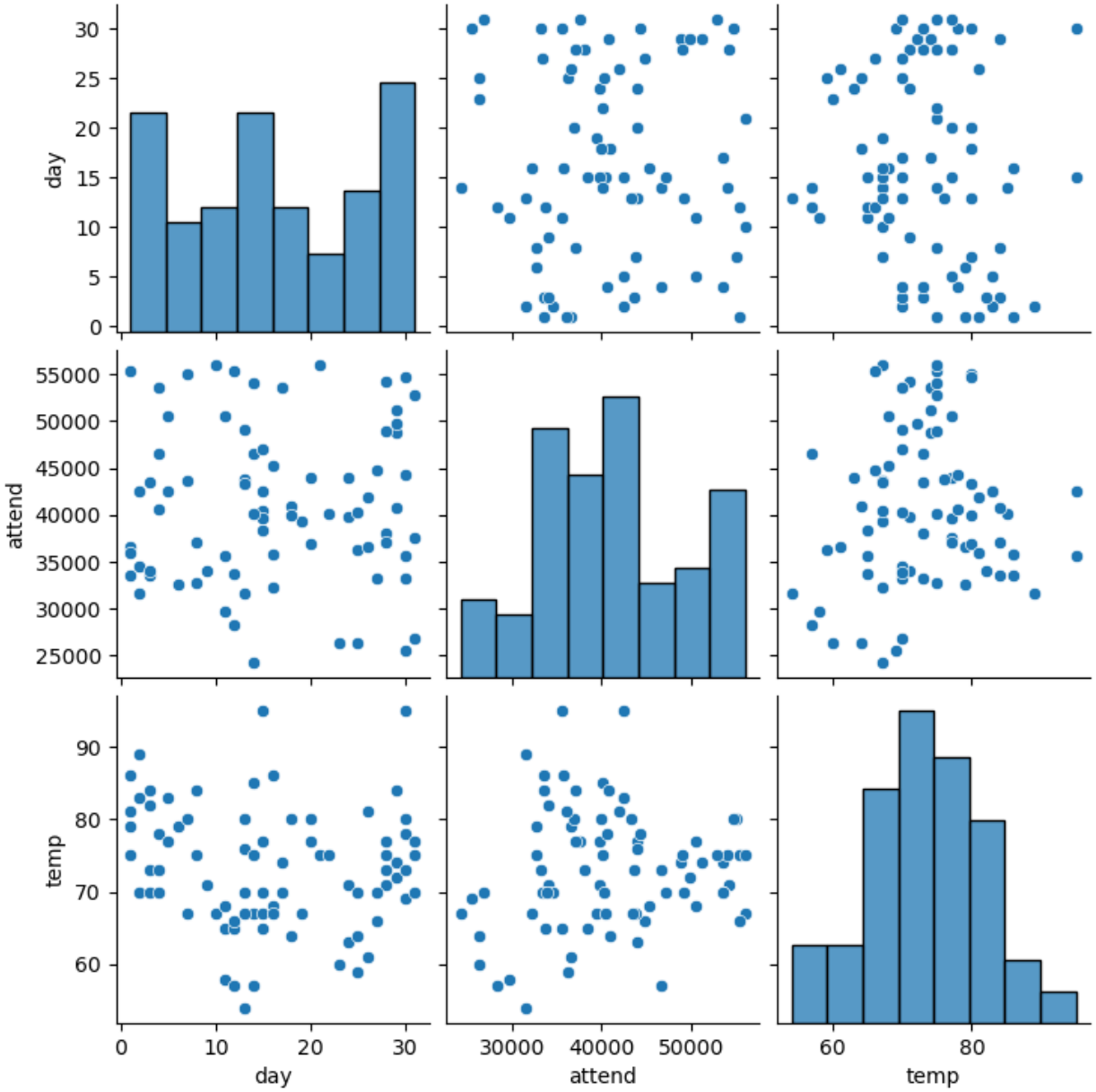
	month	day	attend	day_of_week	opponent	temp	skies	day_night	cap	shirt	fireworks
0	APR	10	56000	Tuesday	Pirates	67	Clear	Day	NO	NO	NO
1	APR	11	29729	Wednesday	Pirates	58	Cloudy	Night	NO	NO	NO
2	APR	12	28328	Thursday	Pirates	57	Cloudy	Night	NO	NO	NO
3	APR	13	31601	Friday	Padres	54	Cloudy	Night	NO	NO	YES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

In [29]:
```
#Correlation Matrix
dat.corr()
#Shows a positive relationship between temperature and attendance, a p
#the year and attendance appears to be a strong negative correlation,
```

Out[29]:

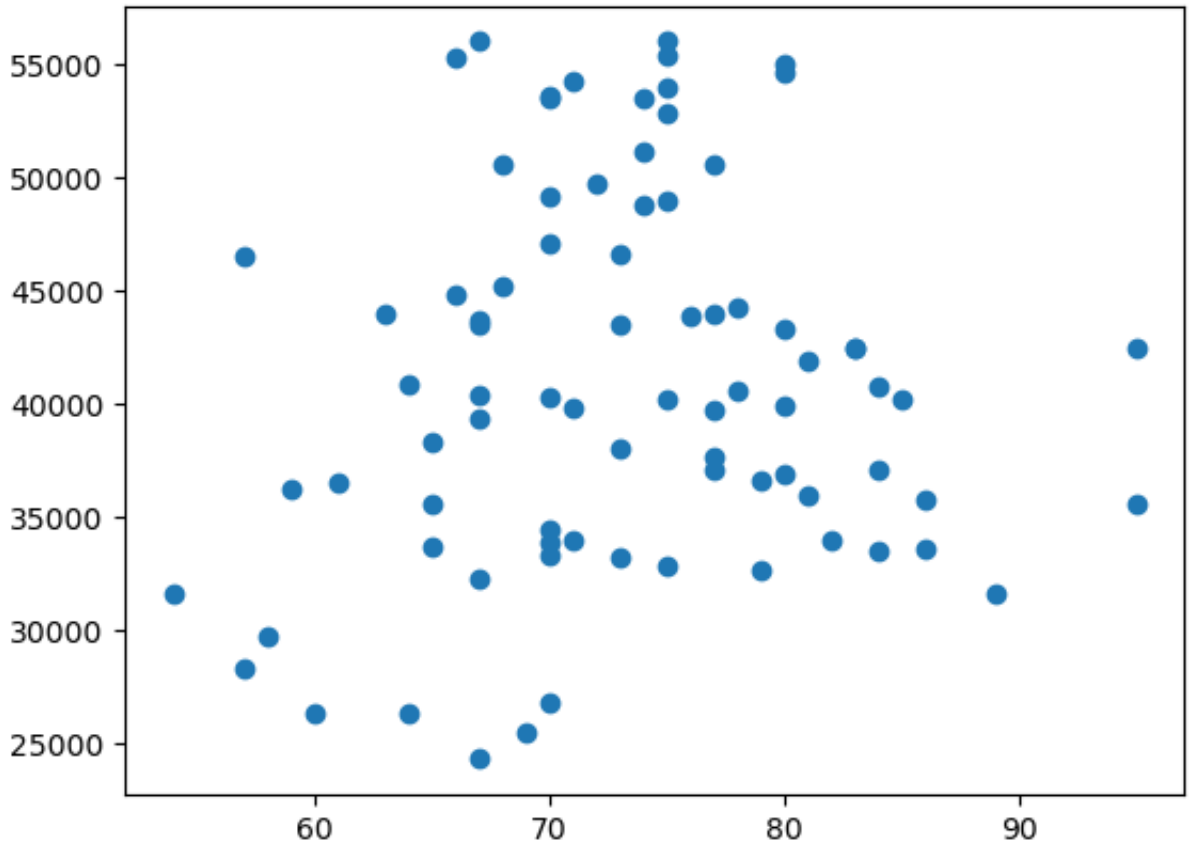
	day	attend	temp
day	1.000000	0.027093	-0.127612
attend	0.027093	1.000000	0.098951
temp	-0.127612	0.098951	1.000000

In [30]:
```
#Correlation Plot of numerical variables.
df = dat[['day', 'attend', 'temp']]
sns.pairplot(df, kind="scatter")
plt.show()
```



In [7]:
```
#We can imagine the strong positive relationship by a line from bottom
#column.
```

In [31]:
```
#Correlation Test.
plt.scatter(dat['temp'], dat['attend'])
plt.show()
```



In [32]:
```
#Calculate the correlation coefficient
np.corrcoef(dat['temp'], dat['attend'])
```

Out[32]:
```
array([[1.        , 0.09895073],
       [0.09895073, 1.        ]])
```

In [33]:
```
#shows a positive relationship of 0.09895.
```

In [34]:
```
#Linear Regression test
from scipy.stats import linregress
linregress(dat['temp'], dat['attend'])
```

Out[34]:
```
LinregressResult(slope=98.71542724060392, intercept=33819.223377770635, rvalue=0.0989507309062371, pvalue=0.3794693886882514, stderr=111.69040295967339, intercept_stderr=8221.94266396445)
```

In [35]:
```
#The p-value of 0.3794693 is greater than 0.05, we fail to reject the
#significant.  So depending on your parameters, you might not want to
```

In [15]:
```
#Categorical Variable relationships.
pd.crosstab(dat.opponent, dat.attend)
```

Out[15]:

attend	24312	25509	26345	26376	26773	28328	29729	31601	31607	32238	...	53498
opponent												
Angels	0	0	0	0	0	0	0	0	0	0	...	0
Astros	0	0	0	0	0	0	0	0	0	0	...	0
Braves	0	0	1	1	0	0	0	0	0	0	...	0
Brewers	0	1	0	0	1	0	0	0	0	0	...	0
Cardinals	0	0	0	0	0	0	0	0	0	0	...	0
Cubs	0	0	0	0	0	0	0	0	0	0	...	0
Giants	0	0	0	0	0	0	0	0	0	0	...	0
Marlins	0	0	0	0	0	0	0	0	0	0	...	0
Mets	0	0	0	0	0	0	0	0	0	0	...	0
Nationals	0	0	0	0	0	0	0	0	0	0	...	0

In [36]:
```
pd.crosstab(dat.cap, dat.attend)
```

Out[36]:

attend	24312	25509	26345	26376	26773	28328	29729	31601	31607	32238	...	53498	53
cap													
NO	1	1	1	1	1	1	1	1	1	1	...	1	
YES	0	0	0	0	0	0	0	0	0	0	...	0	

2 rows × 80 columns

In [41]:
```
dat1 =pd.get_dummies(data=dat, columns=["skies", "day_night", "cap", "
```

In [43]:
```
dat1.head(5)
```

Out[43]:

	month	day	attend	temp	skies_Clear	skies_Cloudy	day_night_Day	day_night_Night	cap_
0	APR	10	56000	67	1	0	1	0	
1	APR	11	29729	58	0	1	0	1	
2	APR	12	28328	57	0	1	0	1	
3	APR	13	31601	54	0	1	0	1	
4	APR	14	46549	57	0	1	0	1	

5 rows × 40 columns

In [44]:
```
#dat['TypeInt']=(dat[''])
result = dat1.dtypes
print(result)
```

```
month                   object
day                      int64
attend                   int64
temp                     int64
skies_Clear              uint8
skies_Cloudy             uint8
day_night_Day            uint8
day_night_Night          uint8
cap_NO                   uint8
cap_YES                  uint8
shirt_NO                 uint8
shirt_YES                uint8
fireworks_NO             uint8
fireworks_YES            uint8
bobblehead_NO            uint8
bobblehead_YES           uint8
day_of_week_Friday       uint8
day_of_week_Monday       uint8
day_of_week_Saturday     uint8
day_of_week_Sunday       uint8
day_of_week_Thursday     uint8
day_of_week_Tuesday      uint8
day_of_week_Wednesday    uint8
opponent_Angels          uint8
opponent_Astros          uint8
opponent_Braves          uint8
opponent_Brewers         uint8
opponent_Cardinals       uint8
opponent_Cubs            uint8
opponent_Giants          uint8
opponent_Marlins         uint8
opponent_Mets            uint8
opponent_Nationals       uint8
opponent_Padres          uint8
opponent_Phillies        uint8
opponent_Pirates         uint8
opponent_Reds            uint8
opponent_Rockies         uint8
opponent_Snakes          uint8
opponent_White Sox       uint8
dtype: object
```

In [61]:
```
pd.options.display.max_rows = 999
pd.options.display.max_rows
```

Out[61]: 999

In [63]:
```
np.set_printoptions(linewidth=110)
```

In [64]:
```
#Correlation Matrix now that they are numerical with the categorical t
results = dat1.corr()
display(results)
```

	day	attend	temp	skies_Clear	skies_Cloudy	day_night_
day	1.000000	0.027093	-0.127612	0.038396	-0.038396	0.03
attend	0.027093	1.000000	0.098951	0.150963	-0.150963	0.04
temp	-0.127612	0.098951	1.000000	0.316584	-0.316584	0.27
skies_Clear	0.038396	0.150963	0.316584	1.000000	-1.000000	0.18
skies_Cloudy	-0.038396	-0.150963	-0.316584	-1.000000	1.000000	-0.18
day_night_Day	0.039828	0.043544	0.272141	0.188903	-0.188903	1.00
day_night_Night	-0.039828	-0.043544	-0.272141	-0.188903	0.188903	-1.00
cap_NO	0.202274	0.055002	-0.064521	0.099671	-0.099671	-0.12
cap_YES	-0.202274	-0.055002	0.064521	-0.099671	0.099671	0.12
shirt_NO	0.030182	-0.133269	-0.004394	-0.108566	0.108566	-0.07
shirt_YES	-0.030182	0.133269	0.004394	0.108566	-0.108566	0.07
fireworks_NO	-0.099528	-0.002094	0.189899	-0.021880	0.021880	0.21
fireworks_YES	0.099528	0.002094	-0.189899	0.021880	-0.021880	-0.21
bobblehead_NO	-0.145363	-0.581895	-0.049573	-0.049349	0.049349	0.18
bobblehead_YES	0.145363	0.581895	0.049573	0.049349	-0.049349	-0.18
day_of_week_Friday	0.145280	-0.048948	-0.182804	0.003920	-0.003920	-0.20
day_of_week_Monday	-0.118786	-0.307198	-0.015883	0.066828	-0.066828	-0.19
day_of_week_Saturday	0.081861	0.107788	-0.024113	-0.075456	0.075456	-0.12
day_of_week_Sunday	0.021966	0.065153	0.264787	0.242046	-0.242046	0.74
day_of_week_Thursday	0.179059	-0.019679	-0.023216	0.020925	-0.020925	-0.12
day_of_week_Tuesday	-0.094301	0.355316	-0.020044	-0.075456	0.075456	-0.12
day_of_week_Wednesday	-0.162472	-0.174723	-0.007474	-0.179220	0.179220	-0.01

opponent_Angels	-0.084966	0.207796	-0.145872	-0.045712	0.045712	-0.09
opponent_Astros	0.202649	-0.134533	-0.232868	-0.199990	0.199990	0.07
opponent_Braves	0.161561	-0.209171	-0.256594	-0.354268	0.354268	-0.09
opponent_Brewers	0.319078	-0.157030	-0.045446	0.126173	-0.126173	-0.10
opponent_Cardinals	0.009433	-0.006967	0.223057	0.170261	-0.170261	-0.03
opponent_Cubs	-0.249317	0.075310	0.075572	-0.045712	0.045712	0.07
opponent_Giants	-0.214901	-0.074763	0.169508	0.103011	-0.103011	-0.16
opponent_Marlins	0.182105	-0.008912	0.020211	0.108566	-0.108566	0.07
opponent_Mets	0.140011	0.236213	0.051063	0.126173	-0.126173	0.03
opponent_Nationals	0.243737	0.195667	-0.066785	0.108566	-0.108566	0.07
opponent_Padres	-0.202555	0.045111	-0.072872	-0.267828	0.267828	0.03
opponent_Phillies	0.017754	0.020380	-0.019332	-0.045712	0.045712	0.07
opponent_Pirates	-0.105510	-0.071849	-0.296138	-0.199990	0.199990	0.07
opponent_Reds	-0.269861	-0.009301	-0.074693	-0.045712	0.045712	-0.09
opponent_Rockies	-0.005030	-0.060404	0.193270	0.103011	-0.103011	0.03
opponent_Snakes	0.040237	-0.073943	0.155250	0.103011	-0.103011	0.03
opponent_White Sox	-0.002790	0.127046	-0.082602	0.108566	-0.108566	0.07

39 rows × 39 columns

In [50]:
```
dat1.head()
```

Out[50]:

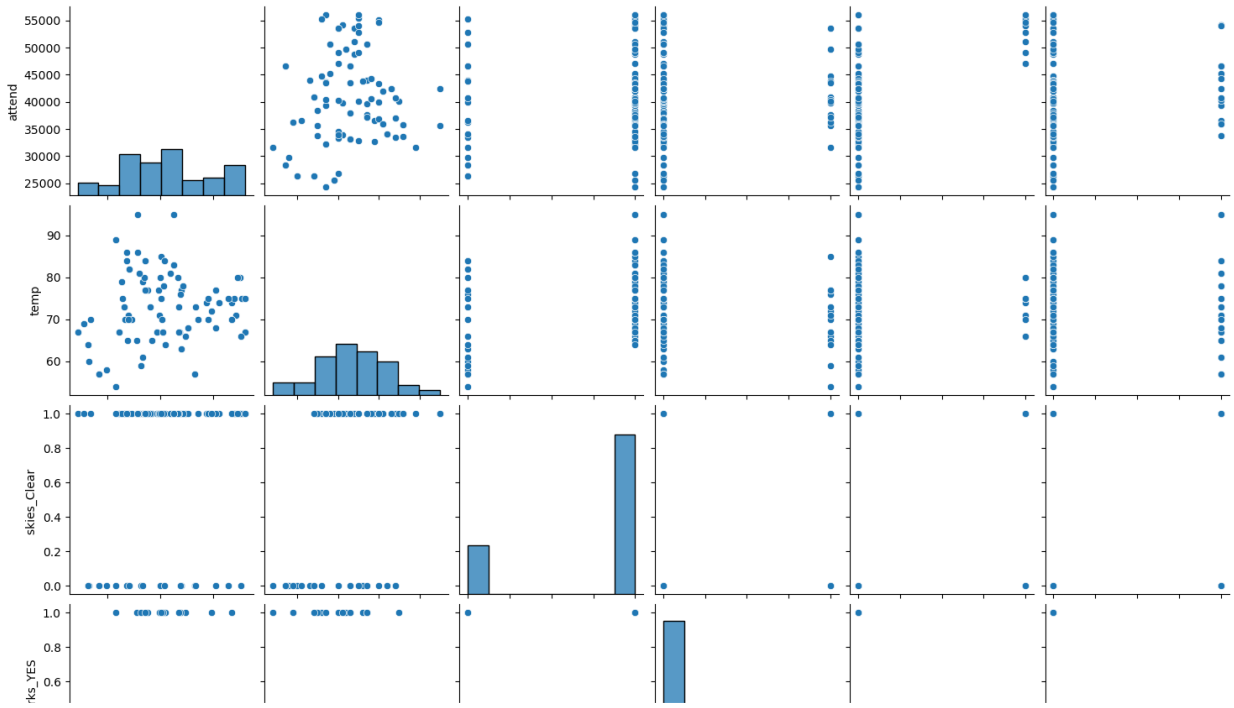
	month	day	attend	temp	skies_Clear	skies_Cloudy	day_night_Day	day_night_Night	cap_
0	APR	10	56000	67	1	0	1	0	
1	APR	11	29729	58	0	1	0	1	
2	APR	12	28328	57	0	1	0	1	
3	APR	13	31601	54	0	1	0	1	
4	APR	14	46549	57	0	1	0	1	

5 rows × 40 columns

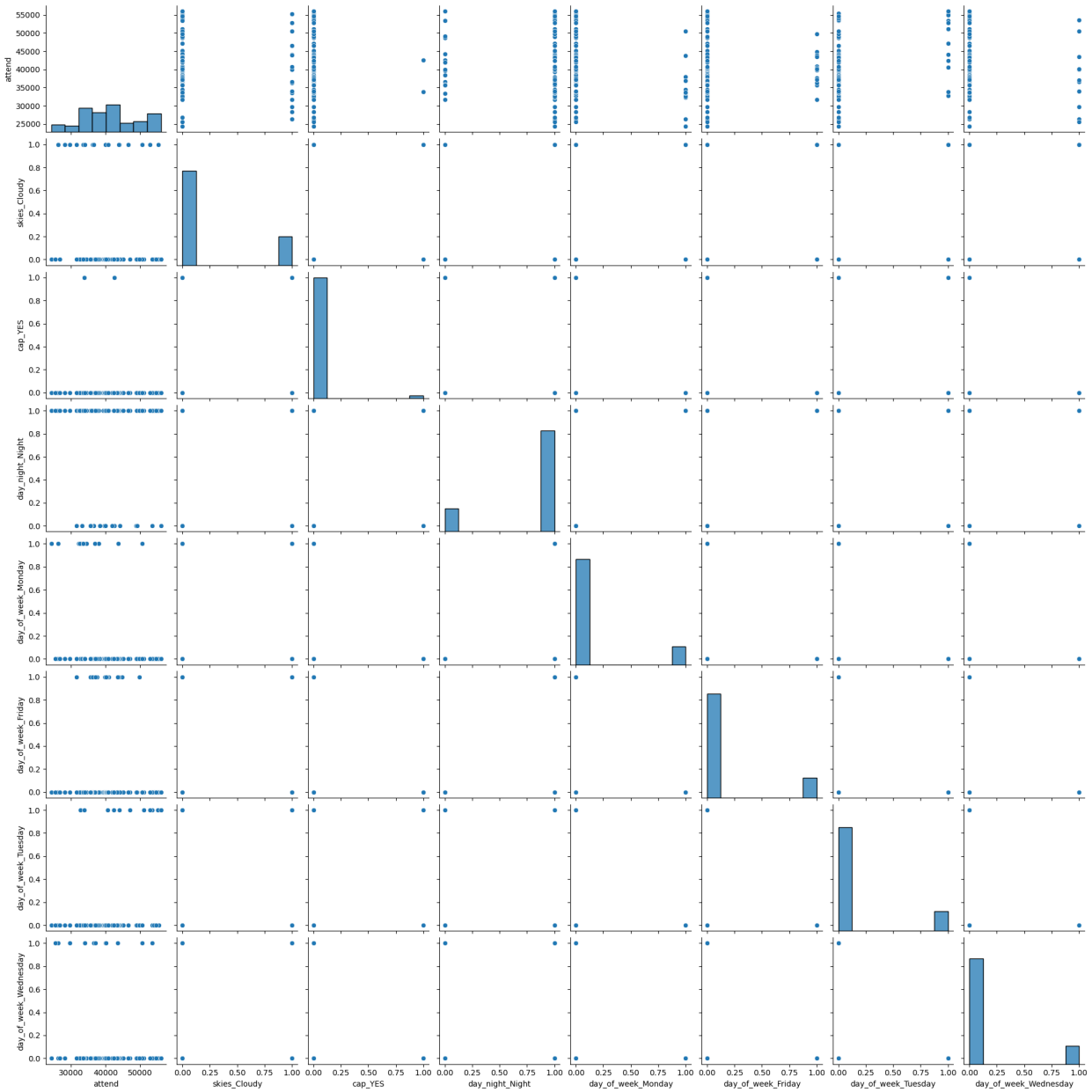
In [52]:
```
list(dat1.columns)
```

Out[52]:
```
['month',
 'day',
 'attend',
 'temp',
 'skies_Clear ',
 'skies_Cloudy',
 'day_night_Day',
 'day_night_Night',
 'cap_NO',
 'cap_YES',
 'shirt_NO',
 'shirt_YES',
 'fireworks_NO',
 'fireworks_YES',
 'bobblehead_NO',
 'bobblehead_YES',
 'day_of_week_Friday',
 'day_of_week_Monday',
 'day_of_week_Saturday',
 'day_of_week_Sunday',
 'day_of_week_Thursday',
 'day_of_week_Tuesday',
 'day_of_week_Wednesday',
 'opponent_Angels',
 'opponent_Astros',
 'opponent_Braves',
 'opponent_Brewers',
 'opponent_Cardinals',
 'opponent_Cubs',
 'opponent_Giants',
 'opponent_Marlins',
 'opponent_Mets',
 'opponent_Nationals',
 'opponent_Padres',
 'opponent_Phillies',
 'opponent_Pirates',
 'opponent_Reds',
 'opponent_Rockies',
 'opponent_Snakes',
 'opponent_White Sox']
```

In [67]:
```
#Correlation Plot of the new numerical variables- positive results cor
#temperature and attendance.
dfc = dat1[[
 'attend',
 'temp',
 'skies_Clear ',
 'fireworks_YES',
 'bobblehead_YES',
 'day_of_week_Saturday']]
sns.pairplot(dfc, kind="scatter")
plt.show()
```



In [66]:
```
#Correlation Plot of the new numerical variables- negative results cor
#on the scatterplots despite negative amounts found in correlation num
dfc = dat1[[
 'attend',
 'skies_Cloudy',
 'cap_YES',
 'day_night_Night',
 'day_of_week_Monday',
 'day_of_week_Friday',
 'day_of_week_Tuesday',
 'day_of_week_Wednesday']]
sns.pairplot(dfc, kind="scatter")
plt.show()
```



In [ ]:
```
#During data analysis only one correlation was observed and it wasn't
#and attendance is the only observed positive relationship.  There was
#a recommendation to enhance attendance would be most suggessful if it
#Promotions like the cap or even fireworks did not show correlation to
#as many home games as possible during higher temperatures to potentia
#recommend less spending on the promotional items including fireworks,
#increasing attendance.
```

In [72]:
```
#!pip install nbconvert
```

In [ ]:
```

```