Answers for Lab 16

Cleaning the data:

1)
```
fields = ['Year Built','Number Of Stories','Land Area (AC)',
'Rentable Building Area','Typical Floor Size','Number Of
Elevators', 'Percent Leased']

ts =
pd.read_csv('Times_Square_Property_Data__Commercial_and_Retail_pro
perties_.csv',usecols =fields)
```

2)
```
ts.columns =
['Year','Stories','Land_area','Rentable_area','Floor_size','Elevat
ors','Percent_leased']
```

3)
In interactive shell:
```
ts.head()
```

4) `ts = ts.dropna(axis=0, how = 'any')`

Exploratory Data Analysis:

1) `corr = ts.corr()`

2) At top: `import seaborn as sns`
At end of current code: `sns.heatmap(corr)`

3) `plt.show()`   `# print and clear previous plot`
`sns.pairplot(ts)`

Model Building and Prediction:
1) At top: `import statsmodels.formula.api as smf`
   At end of current code:
   `lm = smf.ols(formula = 'Stories ~ Elevators', data =`
`ts).fit()`

2) `print(lm.params)`

This prints:
```
Intercept    6.966129
Elevators    1.518004
```

The number after `Elevators` is its coefficient, and can be interpreted as the model saying that every new elevator corresponds to 1.518 new stories.

3)
```
plt.show() # print and clear previous plot
sns.regplot(x = 'Stories', y = 'Elevators')
```

4)
```
# make a new DataFrame which is a single column (Elevators)
# with 1 row (20)
elev_new = pd.DataFrame({'Elevators':[20]})
```