Assume the first 5 rows of the dataset will be provided.

1) For the Titanic dataset, how would you test the hypothesis that passengers 10 or younger were more likely to survive than passengers older than 10? Describe the steps in English, not using code.

2) Consider the following dataframe, stored in the variable `df`:

Name	Num	Col1
NY	2	7.4
NJ	5	-3
CT	1	2
PA	3	19.2

Assume each piece of code is in a separate program, so they do not depend on each other.

a)

```
f = df['Col1'] < 5
new_df = df[f]
new_df['Col2'] = new_df['Num']*2
```

What are the contents of `new_df`?

b) What does `print(len(df))` output?

c) What does this code print?

```
f2 = (df['Col1'] > 0) & (df['Num'] > 2)
new_df = df[f2]
m = new_df['Col1'].mean()
print("Mean is",m)
```

d) Draw the graph:

```
df.plot.scatter(x = 'Num', y = 'Col1', title = "My Graph")
```

3) Using the Times Square Building dataset:
https://data.cityofnewyork.us/City-Government/Times-Square-Property-Data-Commercial-and-Retail-p/j86k-5i43

a) Write Python code to estimate the probability that a building in Times Square has 20 or more stories.

b) Write Python code to compute the average typical floor size in buildings with 10-20 elevators.

c) Write Python code to plot a histogram of the rentable building area in buildings with 10 or fewer stories. The histogram should have a title and 15 bins.

d) Write Python code to count the number of buildings with 0 elevators and 5 or more stories.

4) Using the Times Square Building datatset, write Python code to do the following:

a)Sample 50 buildings, and compute the mean land area for the buildings in the sample.
b) Repeat step (a) 5000 times, storing the mean land area for each sample.
c) Plot a histogram of these stored mean land areas.
d) Plot a boxplot of these stored mean land areas.
e) What do we expect the distribution of these sample mean land areas to look like?