

Assignment 6.2

Ans 1

```
import matplotlib.pyplot as plt

titanic = pd.read_csv('../_datasets/titanic.csv')

# Display the box plots on 3 separate rows and 1 column
fig, axes = plt.subplots(nrows=3, ncols=1)

# Generate a box plot of the fare prices for the First passenger class
titanic.loc[titanic['pclass'] == 1].plot(ax=axes[0], y='fare', kind='box')

# Generate a box plot of the fare prices for the Second passenger class
titanic.loc[titanic['pclass'] == 2].plot(ax=axes[1], y='fare', kind='box')

# Generate a box plot of the fare prices for the Third passenger class
# Display the plot
titanic.loc[titanic['pclass'] == 3].plot(ax=axes[2], y='fare', kind='box')

plt.show()
```

Ans 2

```
titanic_df.head()
and notice that variable "Sex", "Pclass", and "Embarked" are categorical.
Fourth, convert categorical into dummies:
# "sex" is coded as 1=female and 0=male

def convert_sex(sex):
    if sex == 'female':
        sex = 1
    else:
        sex = 0
    return sex
titan_df = titanic_df.copy()
titan_df['Sex'] = titan_df['Sex'].apply(convert_sex)
titan_df.head()
# "Pclass", "Embarked" are categorical, so we get their dummies for analysis

class_dummies = pd.get_dummies(titan_df['Pclass'], prefix='class')
embark_dummies = pd.get_dummies(titan_df['Embarked'], prefix='port')

# and concat the data with columns of dummies
```

```
titan_df = pd.concat([titan_df,class_dummies,embark_dummies],axis=1)
```

Finally, check the descriptive statistics of all variables, and notice any variables that have missing values,

```
titan_df.describe()
```

```
titan_df['Embarked'].count()
```