Directions:

- You may use your notes and online resources in Moodle or on the web.

- Download the Jupyter notebook **T1_II.ipynb** and rename it **username.ipynb** where **username** is your Rose-Hulman user name.

- Download the data files **wine.csv, titanic.csv, vehicle.csv** from Moodle.

- Enter your computer code between the comment lines labeled **# ENTER CODE HERE**.

- When you are finished with the test, upload your Jupyter notebook to the Test 1 dropbox in Moodle in the following **two** formats: `.ipynb`, `.html`

- The readability and simplicity of your Pandas code will be considered in determining your grade.

- Failure to follow instructions will result in a loss of credit.

1. **Box Plot** (5 pts) The data set **wine.csv** contains data on the chemical properties of various bottles of wine and their corresponding taste quality rating. The first five data records of the **wine.csv** data set are shown below. Use Pandas to create a box plot of the taste quality ratings.

	density	pH	alcohol	quality
0	1.0010	3.00	8.8	6.0
1	0.9940	3.30	NaN	6.0
2	0.9951	3.26	10.1	6.0
3	NaN	NaN	NaN	NaN
4	0.9956	3.19	9.9	6.0

2. **Cleaning** (5 pts) The data set **wine.csv** shown above contains missing values. Use Pandas to do the following:

   (i) First, remove data records that are missing all of their attribute values.

   (ii) Then, replace any missing attribute values by their attribute averages.

3. **Grouping and Plotting** (5 pts) The data set **titanic.csv** contains information on the passengers and crew members on the Titanic when it sank on April 14, 1912.

	Class	Sex	Age	Survived	Freq
0	1st	Male	Child	No	0
1	2nd	Male	Child	No	0
2	3rd	Male	Child	No	35
3	Crew	Male	Child	No	0
4	1st	Female	Child	No	0

   Use Pandas to construct the bar chart shown below describing the probability of surviving the sinking of the Titanic.

   (Continued on the back side.)
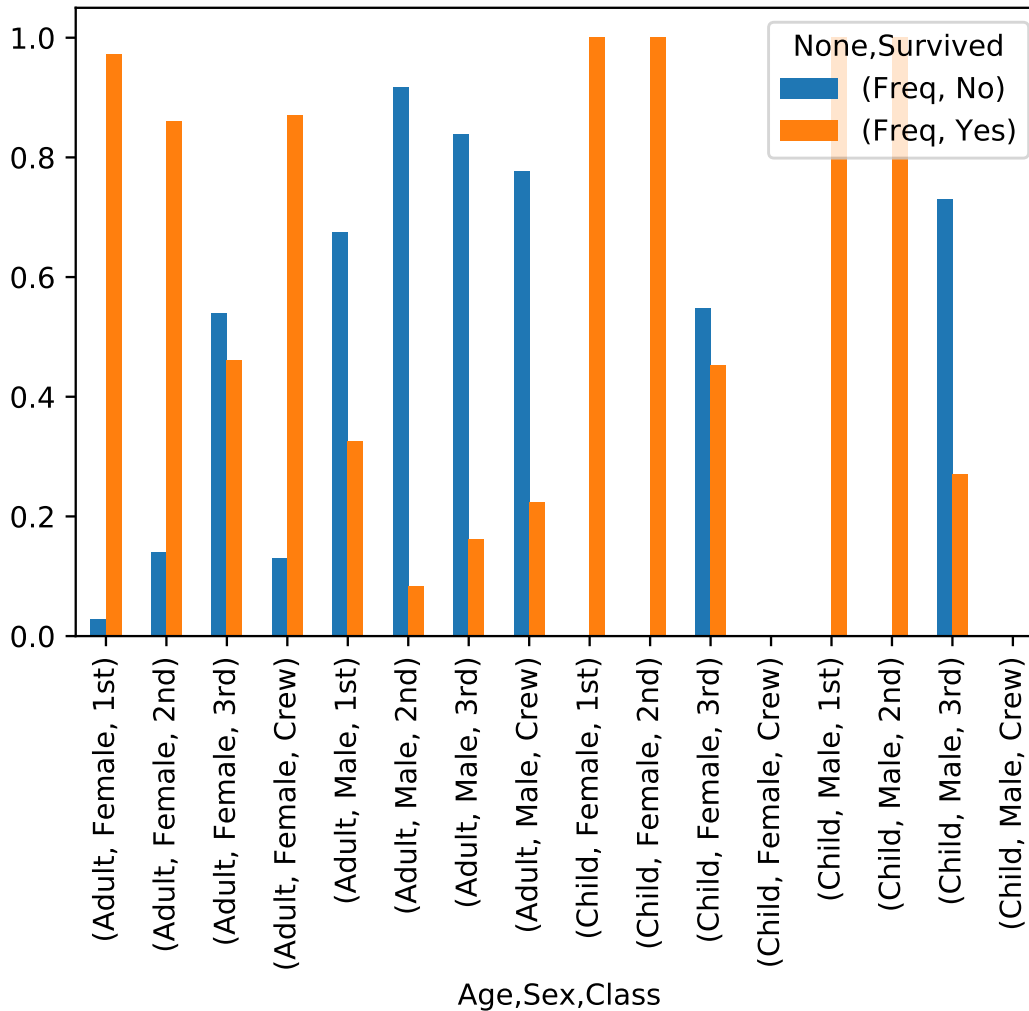
4. **Grouping and Merging** (5 pts) The data set **vehicle.csv** contains data on the combined city & highway fuel economy in miles per gallon (`Cmb MPG`) of various automobiles.

	Model	Displ	Cyl	Drive	Veh Class	Cmb MPG
0	ACURA MDX	3.7	6.0	4WD	SUV	18.0
1	ACURA MDX	3.7	6.0	4WD	SUV	18.0
2	ACURA RDX	2.3	4.0	4WD	SUV	19.0
3	ACURA RDX	2.3	4.0	4WD	SUV	19.0
4	ACURA RDX	2.3	4.0	2WD	SUV	21.0

Use Pandas to construct the expanded table shown below which includes statistics on fuel economy for each class of vehicle (`Veh Class`). Note that automobiles are sorted by fuel economy.

<u>Hint</u>: Use the `.describe` method.

	Model	Displ	Cyl	Drive	Veh Class	Cmb MPG	count	mean	std	min	25%	50%	75%	max
2072	TOYOTA Prius	1.8	4.0	2WD	midsize car	50.0	261.0	24.42	5.93	13.0	20.0	24.0	28.0	50.0
2076	TOYOTA Prius c	1.5	4.0	2WD	small car	50.0	824.0	23.46	5.81	10.0	20.0	22.0	27.0	50.0
2077	TOYOTA Prius c	1.5	4.0	2WD	small car	50.0	824.0	23.46	5.81	10.0	20.0	22.0	27.0	50.0
2073	TOYOTA Prius	1.8	4.0	2WD	midsize car	50.0	261.0	24.42	5.93	13.0	20.0	24.0	28.0	50.0
914	HONDA Civic Hybrid	1.5	4.0	2WD	small car	44.0	824.0	23.46	5.81	10.0	20.0	22.0	27.0	50.0