Directions:

- The time limit for this test is 60 minutes.
- You may use your notes and online resources in Moodle or on the web.
- Download the Jupyter notebook **T1B.ipynb** and rename it **username.ipynb** where **username** is your Rose-Hulman user name.
- Download the data files **restaurant.csv**, **seats.csv**, and **demographics.csv** from Moodle.
- Enter your computer code between the comment lines labeled **# ENTER CODE HERE**.
- When you are finished, execute your Jupyter notebook from beginning to end so that cells are numbered consecutively beginning with cell number 1.
- Upload your Jupyter notebook to the Test 1B dropbox in Moodle in the following **two** formats: `.ipynb` and `.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. (10 pts) **Cleaning:** Load the file `restaurant.csv`. The first three data records are shown below.

	total_bill	tip	sex	smoker	day	time	size
0	25.89	5.16	Male	Yes	Sat	Dinner	4
1	23.68	NaN	Male	No	Sun	Dinner	2
2	12.60	1.00	Male	Yes	Sat	Dinner	2

   - Drop data records with any missing attribute values.
   - Replace all tip values equal to -999 with a value of 0.
   - Your final answer should be named df1. (If you need to, use the method `.copy()` to create a copy of your dataframe.)

2. Load the file `seats.csv`. The first three data records are shown below.

	Price	Advertising	ShelveLoc	Urban	US	Sales
0	120	11	Bad	Yes	Yes	9.50
1	83	16	Good	Yes	Yes	11.22
2	80	10	Medium	Yes	Yes	10.06

   The dataset `seats.csv` contains data on the sales (in thousands of dollars) of car seats in various stores.

   (a) (10 pts) **Binning:** The `Price` attribute is the price of the car seat and `Advertsing` is the local advertising budget (in thousands of dollars).
      - Bin the `Price` attribute so that the bottom 25% of prices are labeled `low`, the top 25% of prices are labeled `high` and the remaining prices are labeled `medium`.
      - Bin the `Advertising` attribute so that values greater than or equal to 0 but less than 1 are labeled `low`, values greater than or equal to 1 but less than 15 are labeled `medium` and values greater than or equal to 15 but less than 30 are labeled `high`.
      - Your final answer should be named df2a.

   (Continued on next page.)

(b) (10 pts) **Merging:** Load the file `demographics.csv`. The data records are shown below.

	Urban	US	Income	Population	Age
0	No	Yes	71.0	285.0	52.0
1	Yes	No	67.0	247.0	53.0
2	Yes	Yes	70.0	266.0	54.0

- Merge demographics data with the original seats data loaded above. No information should be lost in the merging process.
- Your final answer should be named df2b.

	Price	Advertising	ShelveLoc	Urban	US	Sales	Income	Population	Age
0	120	11	Bad	Yes	Yes	9.50	70.0	266.0	54.0
1	83	16	Good	Yes	Yes	11.22	70.0	266.0	54.0
2	80	10	Medium	Yes	Yes	10.06	70.0	266.0	54.0

(c) (10 pts) **Grouping and Plotting:** Constuct the bar chart shown below which represents the average sales in stores based on locations in urban vs non-urban and US vs non-US locations and the shelve locations of the car seats in each store. The order of the bars is important.