

Homework Assignment 1 Cheat Sheet – Exploratory Data Analysis

Dr. Gail Gilboa-Freedman
Dr. Naveh Eskinazi

PYTHON CODING**LIBRARY: PANDAS**

Purpose	Command
Import Library	<code>import pandas as pd</code>
Assign the CSV file into a variable	<code>datafile = 'filename.csv'</code>
Read CSV file into a dataset	<code>dataset = pd.read_csv(datafile)</code>
Assign the dataset into a dataframe	<code>dataframe = pd.DataFrame(dataset)</code>
Show the dataframe content	<code>dataframe</code>
Number of unique values in a column	<code>unique = len(pd.unique(dataframe['column_name']))</code>
Show the data of range of rows	<code>dataframe.iloc[index1:index2]</code>
Create a list of cases that meet a criterion (result: true / false)	<code>index = dataframe['column_name'] == criteria</code>
Show the actual data of the cases	<code>dataframe[index]</code>
Show the descriptive statistics of all numerical columns within the dataframe	<code>dataframe.describe()</code>
Show the number of cases for each unique value in a column	<code>dataframe['column_name'].value_counts()</code>

LIBRARY: MATPLOTLIB

Purpose	Command
Import Library	<i>Import matplotlib as plt</i>
Import pyplot module	<i>Import matplotlib.pyplot as plt</i>
Visualize data of a column using Box & Whisker	<i>dataframe.boxpot(column='column_name')</i>
Create a histogram of a given column with bins defined as column's number of unique values	<i>plt.hist(dataframe['column_name'], bins = unique)</i>
Define x-axis data	<i>x = dataframe ['column_name']</i>
Define y-axis data	<i>y = dataframe ['column_name']</i>
Create a bar chart	<i>plt.bar(x, y)</i>
Define x-axis title	<i>plt.xlabel('enter your text')</i>
Define y-axis title	<i>plt.ylabel('enter your text')</i>
Define chart title	<i>plt.title('enter your text')</i>

Good Luck!