I honestly should have made a docker container for this?

In order to run this code you need:

To be on a Unix system linux mint(debian based should work) is the one I used:

java 8 (you might need to install this)

Python 2.7

pyspark

pandas

tk

matplotlib.

I have inculded a script that should install the libraries you need (excluding python and PIP). You will need to install python and PIP on your own and these need to be installed first before you run the script. (I didn't include it because the installation can be different for your OS) The script should work on linux mint without issues, just make sure to run as sudo. Once you install python and pip you can run the script with the command: ./setup.sh

If the script does not work you will need to do these steps in order:

Update your package manager: The following steps might not work if its not updated

Install python: you need python 2.7

Install PIP: you need the python 2 version of pip

https://www.tecmint.com/install-pip-in-linux/

Install java8: pyspark uses java8 you may need to set your default version of java to java 8 if a different version of java is already installed -> sudo update-alternatives --config java

install -> apt-get install openjdk-8-jdk-headless -qq

switch version -> https://aboullaite.me/switching-between-java-versions-on-ubuntu-linux/

Install python-tk: just get the latest version. I believe this is for graphics

Install tk: if the above does not work get this ? sudo apt install tk

Install Pyspark: python2 -m pip install pyspark

Install Pandas: python2 -m pip install pandas

Install matplotlib: python2 -m pip install matplotlib

To run:

1. put the BeerDataScienceProject.csv into the directory where you downloaded the project files

2. Next use the command python2 data.py

You should be able to run this on windows but you will have to install the libraries on your own as well.

After everything is installed, while you are in the working directory just type python2 data.py which will run the first question.

You will need to uncomment the other questions at the bottom in order to run them.

The answers and descriptions are in the folders named questionX.

Email me if you have issues I can walk you through everything:

ln.jrm.1988@gmail.com