jupyter\_cheatsheet.docx

Contents

[jupyter\_cheatsheet.docx 1](#_Toc19365417)

[Compress a Directory for Download and Transfer out of Notebook 2](#_Toc19365418)

[Execute OS Commands 2](#_Toc19365419)

[Force plots to be inline 3](#_Toc19365420)

[Open iPython notebook 3](#_Toc19365421)

[Open jupyter notebook 3](#_Toc19365422)

[Open Multiple Notebooks Simultaneously 4](#_Toc19365423)

[Querying the Underlying Hadoop System 4](#_Toc19365424)

[Display Detailed Hadoop Memory Usage 4](#_Toc19365425)

[Display Total Hadoop Data Storage for a User 4](#_Toc19365426)

[Read Parquet Data Files 4](#_Toc19365427)

[Troubleshooting 5](#_Toc19365428)

[Unable to Save and Checkpoint; ‘\_xsrf’ argument missing from POST advisory 5](#_Toc19365429)

[Using jupyter notebok with conda 5](#_Toc19365430)

[Allow use of a conda environment 5](#_Toc19365431)

[Where iPython notebooks are stored 5](#_Toc19365432)

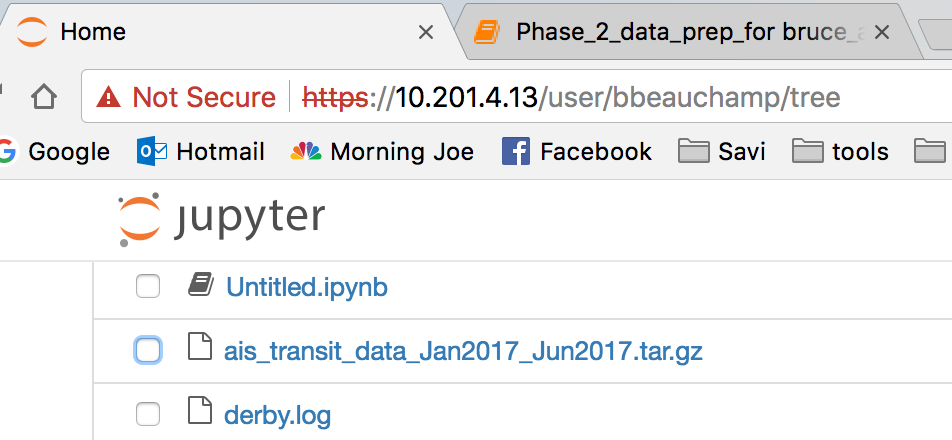
# Compress a Directory for Download and Transfer out of Notebook

# Store the contents of the leg\_models directory to a compressed file

**!tar vcfz** **ais\_transit\_data\_Jan2017\_Jun2017.tar.gz** **leg\_models**

**<compressed output file name----------> <folder to compress>**

Puts the compressed output file into the same directory as the notebook:



# Execute OS Commands

Use ! followed by the underlying OS command, like for unix:

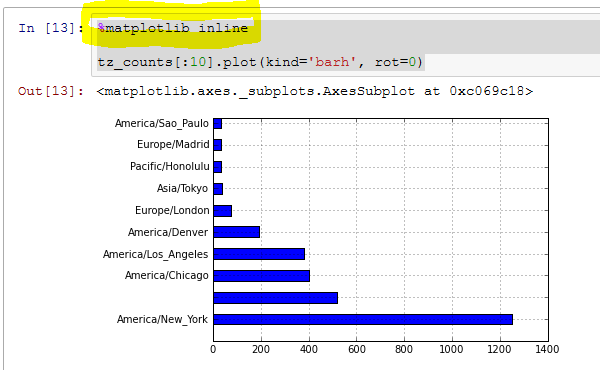
!pwd

output:

/Users/bbeauchamp/Documents/school/reproducible\_data\_analysis

# Force plots to be inline

%matplotlib inline



# Open iPython notebook

C:\Users\Bruce>ipython notebook

[W 08:26:33.769 NotebookApp] ipywidgets package not installed. Widgets are unavailable.

[I 08:26:33.789 NotebookApp] Serving notebooks from local directory: C:\Users\Bruce

[I 08:26:33.789 NotebookApp] 0 active kernels

[I 08:26:33.789 NotebookApp] The IPython Notebook is running at: http://localhost:8888/

[I 08:26:33.789 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to s

kip confirmation).

# Open jupyter notebook

**teja\_code\_review** **jupyter notebook**

[I 09:21:40.182 NotebookApp] Serving notebooks from local directory: /Users/bbeauchamp/Desktop/teja\_code\_review

[I 09:21:40.183 NotebookApp] 0 active kernels

[I 09:21:40.183 NotebookApp] The Jupyter Notebook is running at: http://localhost:8888/?token=ce5496d30e6a95af76ccf4c2bf170f715dbd9b6523ee829f

[I 09:21:40.183 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).

[C 09:21:40.183 NotebookApp]

Copy/paste this URL into your browser when you connect for the first time,

to login with a token:

http://localhost:8888/?token=ce5496d30e6a95af76ccf4c2bf170f715dbd9b6523ee829f

[I 09:21:40.307 NotebookApp] Accepting one-time-token-authenticated connection from ::1

# Open Multiple Notebooks Simultaneously

Create a folder for each new notebook. Since the notebook is in another folder it will create a new metastore\_db per notebook.

# Querying the Underlying Hadoop System

## Display Detailed Hadoop Memory Usage

!hadoop fs -du -h

Output:

1.2 T 1.2 T AIS\_2017.p

189.9 K 189.9 K Syngenta

6.1 K 6.1 K home

6.1 K 6.1 K pg\_lanes\_export\_2017-05-18.csv

99.7 K 99.7 K sample.csv

55.4 K 55.4 K shipment\_modes\_by\_carr\_corr.csv

189.9 K 189.9 K syngenta\_trailer\_daily\_km.csv

104.0 M 104.0 M trip\_duration\_imo.csv

## Display Total Hadoop Data Storage for a User

!hadoop fs -du -s -h /user/bbeauchamp

Output:

1.2 T 1.2 T /user/bbeauchamp

# Read Parquet Data Files

# Try reading all the POC shipment summaries.

shipment\_summary = sqlContext.read.parquet('/lambda/summaries-parquet/POC/shipment-summaries//quarter=2018-q1/\*',

'/lambda/summaries-parquet/POC/shipment-summaries//quarter=2018-q2/\*',

'/lambda/summaries-parquet/POC/shipment-summaries//quarter=2018-q3/\*')

shipment\_summary.registerTempTable('shipment\_sums')

test\_result = sqlContext.sql(

'''SELECT \*

FROM shipment\_sums

LIMIT 1

''')

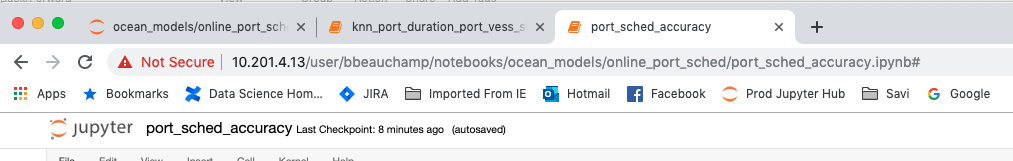
test\_result.count()

test\_result.show()

# Troubleshooting

## Unable to Save and Checkpoint; ‘\_xsrf’ argument missing from POST advisory

1. Refresh using the browser:



1. Then save and checkpoint.

# Using jupyter notebok with conda

## Allow use of a conda environment

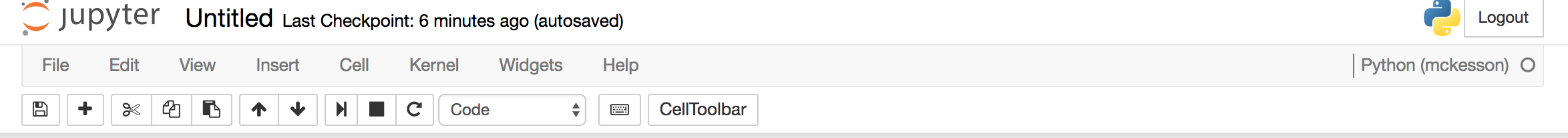
source activate myenv

python -m ipykernel install --user --name myenv --display-name "Python (myenv)"

source activate other-env

python -m ipykernel install --user --name other-env --display-name "Python (other-env)"

This allows the use of a specific environment inside of jupyter notebooks, usable like:



# Where iPython notebooks are stored

C:\Users\Bruce