Team 3 Section 1

Recreate our work and execute R and Python from the same Jupyter Notebook

1. Launch an Amazon EC2 Ubuntu 64x instance (sizedepends on your requirement) and SSH into the instance

```
You can use Putty if you are working on Windows
```

2. Download and install Anaconda

```
wget http://repo.continuum.io/archive/Anaconda3-4.1.1-Linux-x86_64.sh
bash Anaconda3-4.1.1-Linux-x86_64.sh
```

3. Run the below command to ensure that Anaconda's Pythonversion is used

```
source .bashrc
```

4. Configure Jupyter notebook

```
jupyter notebook --generate-config
```

5. Create Certifications

```
mkdir certs
cd certs
sudo openssl req -x509 -nodes -days 365 -newkey rsa:1024 -keyout mycert.pem -out ycert.pem
```

6. Edit Jupyter Configuration file

```
cd ~/.jupyter/
vi jupyter_notebook_config.py
```

7. Once you execute the above vi command, you shouldsee a some commented Python code. Press i on your keyboard to activate-INSERT-. Then at the top of the file type the following code and then save thefile(using :wq)

```
c = get_config()
# Notebook config this is where you saved your pem cert
c.NotebookApp.certfile = u'/home/ubuntu/certs/mycert.pem'
# Run on all IP addresses of your instance
c.NotebookApp.ip = '*'
# Don't open browser by default
c.NotebookApp.open_browser = False
# Fix port to 8888
c.NotebookApp.port = 8888
```

8. Installing Java

```
sudo apt-get update
sudo apt-get install default-jre
```

9. Installing py4j

```
export PATH=$PATH:$HOME/anaconda3/bin
conda install pip
pip install py4j
```

10. Install Spark and Hadoop

```
wget http://archive.apache.org/dist/spark/spark-2.0.0/spark-2.0.0-bin-hadoop2.7.tgz
sudo tar -zxvf spark-2.0.0-bin-hadoop2.7.tgz
```

11. Setting Spark's path

```
export SPARK_HOME='/home/ubuntu/spark-2.0.0-bin-hadoop2.7'
export PATH=$SPARK_HOME:$PATH
export PYTHONPATH=$SPARK_HOME/python:$PYTHONPATH
```

12. Installing R

```
sudo apt-get install r-base r-base-dev libssl-dev libcurl3-dev curl
```

13. Run the following commands inside R terminal

```
install.packages(c('pbdZMQ', 'repr', 'devtools'))
devtools::install_github('IRkernel/IRkernel')
IRkernel::installspec()
```

14. Installing rpy2

```
conda install -c rmkoesterer rpy2
```

15. Installing all essential R packages

```
conda install r-essentials
```

16. Launching Jupyter Notebook

```
jupyter notebook
```

You are good to go!