Spark & MySQL

remember to do
```

brew services list
brew services start xxx
brew services stop xxx
````


Starting mysql server locally (not automatically when booting, but on demand)
mysql.server start
mysql -u root. (from brew installation)
mysql.server stop

create user 'ckapsalis'@'localhost' identified by 'ckapsalis';
grant all privileges on *.* to 'ckapsalis'@'localhost';
show databases;

!!! We need to use JDBC (Java Database Connectivity) to connect Spark to our MySQL database and load a table as a Spark DataFrame.
I downloaded it from
https://dev.mysql.com/downloads/connector/j/
(downloaded the platform agnostic version)
Moved the unzipped folder to /Users/chkapsalis/Documents/mysql-connector-j-9.2.0

****
pip install pyspark
pip install findspark

(base) ChristoorossAir:~ chkapsalis$ pyspark
/Users/chkapsalis/.pyenv/versions/base/lib/python3.10/site-packages/pyspark/bin/spark-class: line 71: /usr/lib/jvm/java-11-openjdk-amd64/bin/java: No such file or directory
/Users/chkapsalis/.pyenv/versions/base/lib/python3.10/site-packages/pyspark/bin/spark-class: line 97: CMD: bad array subscript
head: illegal line count -- -1

(base) ChristoorossAir:~ chkapsalis$ python -c "import findspark; findspark.init(); print(findspark.find())" /Users/chkapsalis/.pyenv/versions/base/lib/python3.10/site-packages/pyspark

```
export SPARK_HOME=/Users/chkapsalis/.pyenv/versions/base/lib/python3.10/
site-packages/pyspark
echo 'export SPARK_HOME=/Users/chkapsalis/.pyenv/versions/base/lib/
python3.10/site-packages/pyspark' >> ~/.zshrc
echo 'export SPARK_HOME=/Users/chkapsalis/.pyenv/versions/base/lib/
python3.10/site-packages/pyspark' >> ~/.bashrc
source ~/.zshrc
source ~/.bashrc
```

```
(base) ChristoorossAir:~ chkapsalis$ pyspark
/Users/chkapsalis/.pyenv/versions/base/lib/python3.10/site-packages/pyspark/
bin/spark-class: line 71: /usr/lib/jvm/java-11-openjdk-amd64/bin/java: No such
file or directory
/Users/chkapsalis/.pyenv/versions/base/lib/python3.10/site-packages/pyspark/
bin/spark-class: line 97: CMD: bad array subscript
head: illegal line count -- -1
```

I need to make sure that openjdk@11 is install and properly linked

Step 1 - force install java 11
```
brew uninstall --ignore-dependencies openjdk@11
brew install openjdk@11
```

Step 2 - manually link java 11
```
sudo ln -sfn /opt/homebrew/opt/openjdk@11/libexec/openjdk.jdk /Library/Java/
JavaVirtualMachines/openjdk-11.jdk
```

Step 3 - step java 11 as default
```
export JAVA_HOME=$(/usr/libexec/java_home -v 11)
echo 'export JAVA_HOME=$(/usr/libexec/java_home -v 11)' >> ~/.zshrc
echo 'export JAVA_HOME=$(/usr/libexec/java_home -v 11)' >> ~/.bashrc
source ~/.zshrc
source ~/.bashrc
```

Step 4 [if working on a notebook] - make sure that the notebook works on
the same environment as the terminal - installation environment
```
(base) ChristoorossAir:~ chkapsalis$ python -c "import sys;
print(sys.executable)"
/Users/chkapsalis/.pyenv/versions/base/bin/python
```

I need to make sure that this is the kernel used by notebooks
I will create a kernel called "Python (PySpark)" that will be the exact same as
the one i'm working with on my terminal
```
pip install --upgrade ipykernel
```

```
python -m ipykernel install --user --name pyspark --display-name "Python (PySpark)"
```