

```
pip install pyspark
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting pyspark
  Downloading pyspark-3.4.0.tar.gz (310.8 MB)
    310.8/310.8 MB 4.5 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: py4j==0.10.9.7 in /usr/local/lib/python3.10/dist-packages (from pyspark) (0.10.9.7)
Building wheels for collected packages: pyspark
  Building wheel for pyspark (setup.py) ... done
  Created wheel for pyspark: filename=pyspark-3.4.0-py2.py3-none-any.whl size=311317130 sha256=0b661f8a5dbcfa5733cd40f920c2fb5704ce
  Stored in directory: /root/.cache/pip/wheels/7b/1b/4b/3363a1d04368e7ff0d408e57ff57966fcd00583774e761327
Successfully built pyspark
Installing collected packages: pyspark
Successfully installed pyspark-3.4.0
```

```
import pyspark
from pyspark import SparkContext
```

```
sc=SparkContext()
```

```
!apt-get install openjdk-8-jdk-headless -qq > /dev/null
!wget -q http://archive.apache.org/dist/spark/spark-3.1.1/spark-3.1.1-bin-hadoop3.2.tgz
!tar xf spark-3.1.1-bin-hadoop3.2.tgz
!pip install -q findspark
```

```
tar: spark-3.1.1-bin-hadoop3.2.tgz: Cannot open: No such file or directory
tar: Error is not recoverable: exiting now
```

```
import os
os.environ["JAVA_HOME"]="/usr/lib/jvm/java-8-openjdk-amd64"
os.environ["SPARK_HOME"]="/content/spark-3.1.1-bin-hadoop3.2"
```

```
#Build Spark Session
import findspark
findspark.init()
from pyspark.sql import SparkSession
spark = SparkSession.builder.master("local[*]").getOrCreate()
spark.conf.set("spark.sql.repl.eagerEval.enabled", True)
```

```
from google.colab import files
```

```
uploaded = files.upload()
```

```
Choose Files mcdonalds.csv
• mcdonalds.csv(text/csv) - 11522 bytes, last modified: 14/06/2023 - 100% done
Saving mcdonalds.csv to mcdonalds.csv
```

```
mcd_df = spark.read.csv("mcdonalds.csv", header=True, inferSchema=True)
mcd_df
```

item	calories	protien	totalfat	satfat	transfat	cholesterol	carbs	sugar	addedsugar	sodium	menu
McVeggie Burger	402	10.24	13.83	5.34	0.16	2.49	56.54	7.9	4.49	706.13	regular

mcd_df.printSchema()

```

root
|-- item: string (nullable = true)
|-- calories: string (nullable = true)
|-- protien: double (nullable = true)
|-- totalfat: double (nullable = true)
|-- satfat: double (nullable = true)
|-- transfat: double (nullable = true)
|-- cholesterol: double (nullable = true)
|-- carbs: double (nullable = true)
|-- sugar: double (nullable = true)
|-- addedsugar: double (nullable = true)
|-- sodium: double (nullable = true)
|-- menu: string (nullable = true)

```

```

||Chicken Kebab Bur... ||357 ||8.64 ||14.02 ||4.84 ||0.13 ||1.51 ||47.9 ||3.64 ||3.49 ||548.79 ||regular||

```

mcd_df.createOrReplaceTempView("mcd")

```

||Mc Egg Masala burger||290 ||12.45 ||12.27 ||3.04 ||0.11 ||213.09 ||32.89||4.00 ||3.04 ||737.91 ||regular||

```

```

spark.sql('SELECT item, calories, protien, totalfat, menu \
FROM mcd \
WHERE calories >= 600 \
ORDER BY calories DESC')

```

item	calories	protien	totalfat	menu
Chicken Cheese La...	834.36	27.37	45.18	gourmet
Veg Maharaja Mac	832	24.17	37.94	regular
Ghee Rice with Mc...	720	26.91	29.2	regular
Chicken Maharaja...	689	34.0	36.69	regular
Spicy Paneer Wrap	674	20.96	39.1	regular
Cheese Lava Burger	671.06	14.99	33.48	gourmet
McSpicy™ Paneer B...	652	20.29	39.45	regular
Chunky Chipotle A...	641.36	39.47	31.51	gourmet
McSpicy Premium V...	634.71	22.44	39.21	gourmet
McSpicy Premium C...	622.25	31.49	34.65	gourmet

```

spark.sql('SELECT item, sugar, addedsugar, menu \
FROM mcd \
WHERE menu = "breakfast" \
ORDER BY sugar')

```

item	sugar	addedsugar	menu
Hash Brown	0.32	0.0	breakfast
Sausage Mc Muffin	2.38	0.75	breakfast
Egg McMuffin	2.38	0.75	breakfast
Double Cheese McM...	2.59	0.75	breakfast
Sausage Mc Muffin...	2.61	0.75	breakfast
Spicy Egg McMuffin	2.63	1.05	breakfast
Veg McMuffin	3.02	0.75	breakfast
Hot Cake with map...	25.72	13.5	breakfast

```

spark.sql('SELECT item, calories, protien, totalfat, cholesterol, sugar, sodium, menu \
FROM mcd \
WHERE sugar <= 1 AND totalfat <= 10 \
ORDER BY item')

```

item	calories	protien	totalfat	cholesterol	sugar	sodium	menu
4 piece Chicken M...	169	10.03	9.54	24.66	0.72	313.25	regular
Americano (L)	26.71	1.09	0.06	0.55	0.28	0.65	mccafe
Americano (R)	23.07	0.94	0.05	0.48	0.24	0.57	mccafe
Americano (S)	12.87	0.52	0.03	0.27	0.13	0.32	mccafe
Cheese Slice	51.03	3.06	3.99	13.43	0.54	15.0	condiments
Coke Zero Can	0.99	0.0	0.0	0.0	0.0	2.9	beverage
Espresso	12.87	0.52	0.03	0.27	0.13	0.32	mccafe
Hash Brown	140.29	1.93	7.32	0.64	0.32	275.26	breakfast
Medium Wedges	280	5.44	9.79	1.33	0.66	487.76	regular
Piri Piri Mix	17.13	0.51	0.36	0.05	0.66	65.24	condiments
Regular Wedges	204	3.97	7.15	0.97	0.48	356.44	regular
Vedica Natural Mi...	0	0.0	0.0	0.0	0.0	0.0	beverage
L1 Coffee	7	0.0	0.0	0.0	0.0	0.0	regular

[Colab paid products](#) - [Cancel contracts here](#)

✓ 0s completed at 12:27 PM

