Scala, Redis and Bokeh

-Anusha J

Scala

- Developed by Martin Odersky in 2003 at EPFL.
- Scala derived from Scalable Language, Blends object-oriented and functional programming in a statically typed language.
- Runs on JVM and Statically typed.
- Declares variables in two different ways **:** val - immutable, var - mutable
- Basic Types: Byte, Short, Int, Long, Float, ?Double, Char, Boolean
- We have to make distinction between methods and functions. A method is a function that is a member of a class, trait or object.
- For high performance data science projects Scala is best language for Spark.

Object and Class representation

```
object HelloWorld {
  def main(args: Array[String]) {
    println("Hello, World!")
  }
}
```

```
defined object HelloWorld
```

```
class Person{
var name: String = "temp"
var gender: String = "temp"
var age: Int = 0

def walking = println(s"$name is walking")

def talking = println(s"$name is talking")

}
// Creating an object of the Person class
val firstPerson = new Person
```

```
defined class Person
firstPerson: Person = Person@3b9f5be2
```

Redis

- Redis is an open source in-memory data store for use as a database, cache, message broker.
- NoSQL database stores data as key-value pairs.
- It supports various data structures such as strings, hashes, sets, lists, sorted sets, bitmaps, hyperloglogs.
- Very Fast: Can perform around 110,000 SETs per second, about 81,000 GETs per second.
- Redis has client API developed in all popular languages -C, Ruby, Java, JavaScript, Python and more.

Interacting with redis from python

```
import redis
r = redis.Redis(host='127.0.0.1',
                port = 6379,
                decode_responses=True)
```

Storing DataFrame into Redis

```
def storeInRedis(alias, df):
    df_compressed = pa.serialize(df).to_buffer().to_pybytes()
    res = r.set(alias,df_compressed)
    if res == True:
        print(f'{alias} cached')
storeInRedis('data', df)
```

Load DataFrame from Redis

```
def loadFromRedis(alias):
    df = r.get(alias)
    try:
        return pa.deserialize(df)
    except:
        print("No data")
loadFromRedis('diamond_dataset')
```

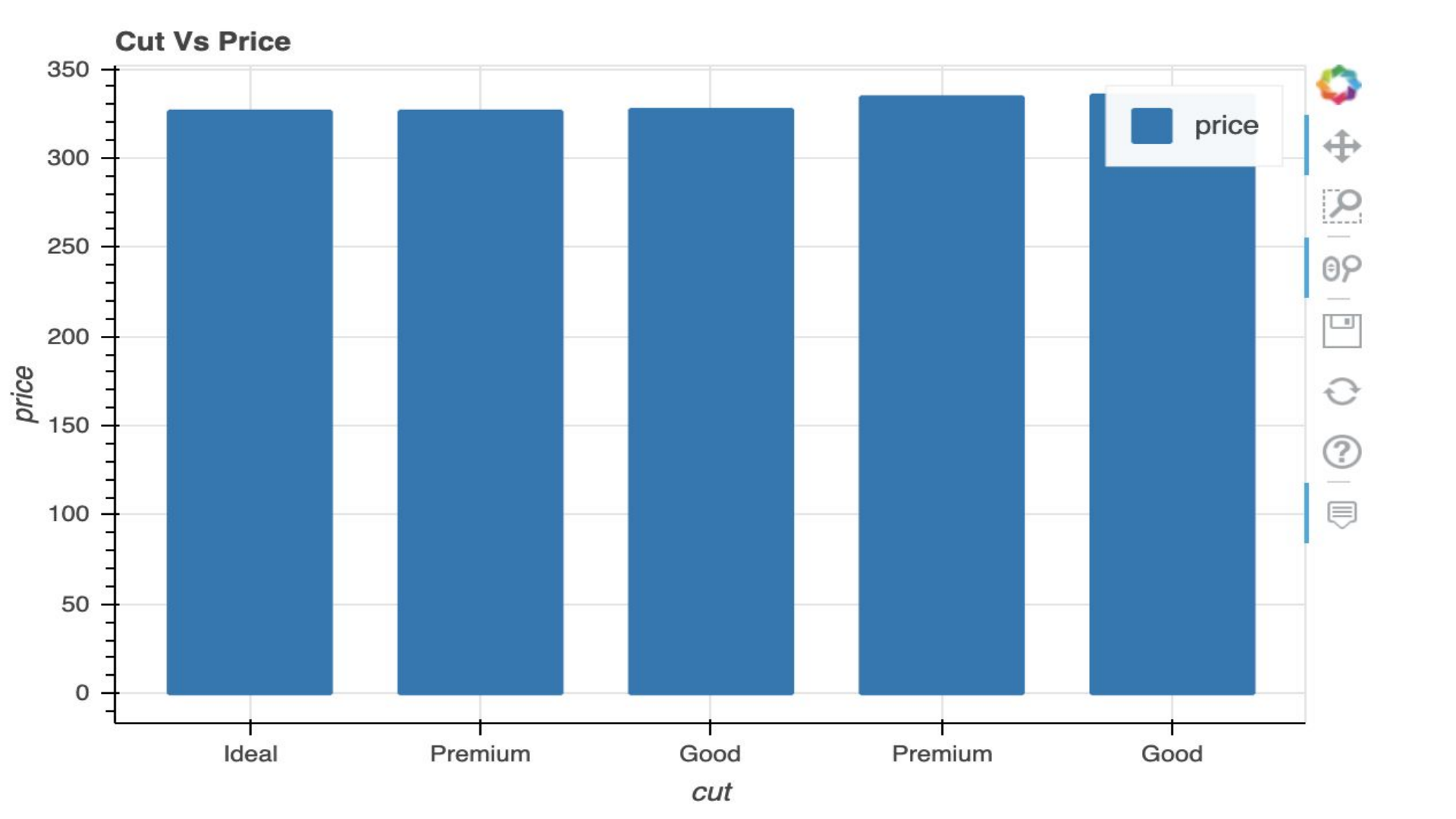
```
storeInRedis('diamond_dataset', df)
```

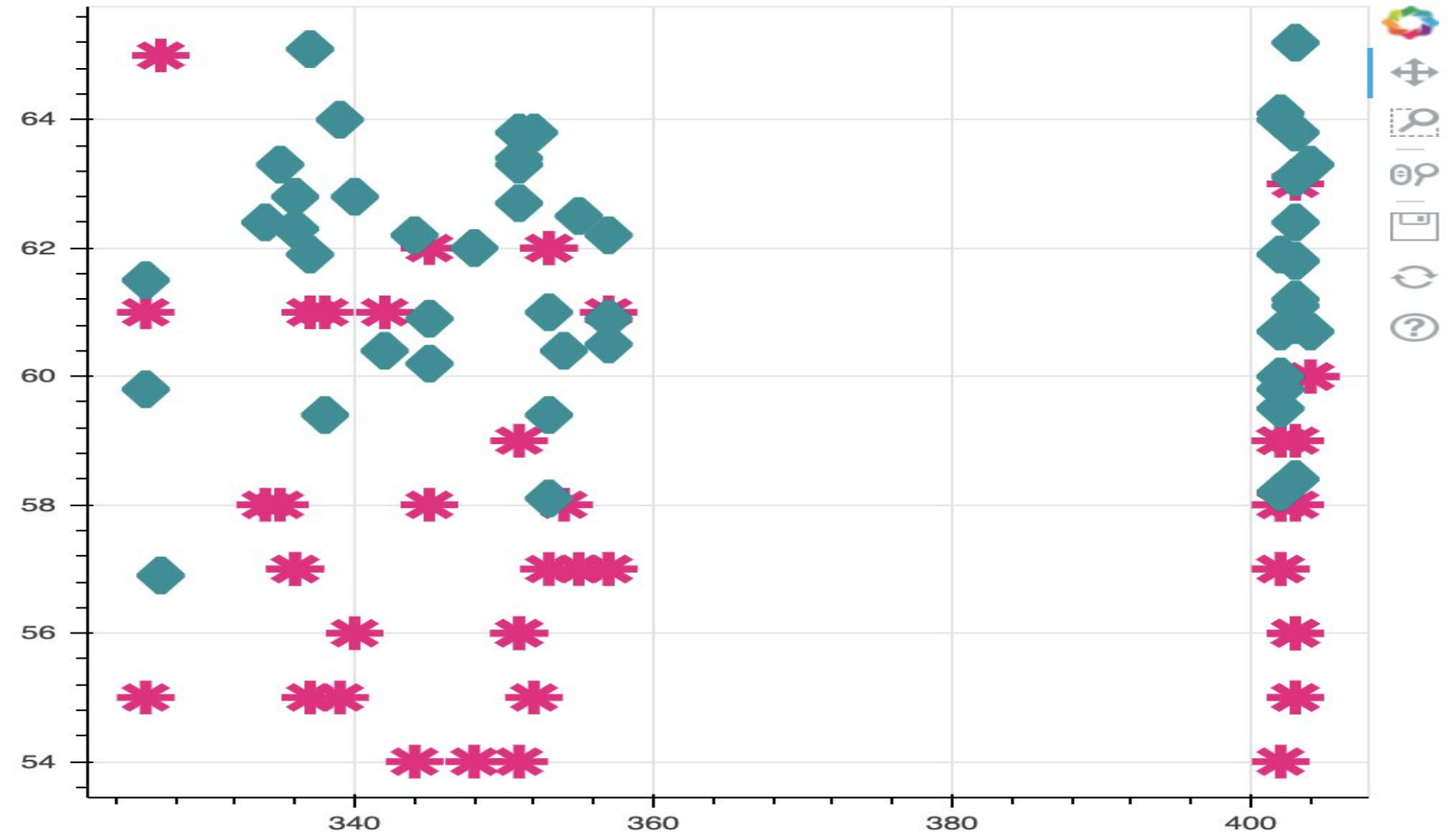
```
diamond_dataset cached
```

```
loadFromRedis('diamond_dataset')
```

Bokeh

- Bokeh is an interactive visualization library, provides elegant, concise construction of versatile graphics.
- Graphs are built up one layer at a time. We start creating a figure, and then we add elements called glyphs, to the figure.
- First, we make a plot using figure method and then we append our glyphs to the plot by passing data.
- To Show our plot in jupyter notebook right below the code should use output_notebook call.
- We get few tools for free with any Bokeh plot which are right side and include zooming, selection, plot saving, refreshing abilities.


Cut Vs Price
price
350
300
250
200
150
100
50
0
price
Ideal
Premium
Good
Premium
Good
cut


64
62
60
58
56
54
340
360
380
400