

RESULTS AND SCREENSHOTS

```
ubuntu@ubuntu:~$ sh /home/ubuntu/Documents/Project/hadoop.sh
-----
-Executing hadoop jobs-
-----
--Starting the cluster--
-----
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [localhost]
ubuntu@localhost's password:
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-ubuntu-namenode-ubuntu.out
ubuntu@localhost's password:
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-ubuntu-datanode-ubuntu.out
Starting secondary namenodes [0.0.0.0]
ubuntu@0.0.0.0's password:
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-ubuntu-secondarynamenode-ubuntu.out
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-ubuntu-resourcemanager-ubuntu.out
ubuntu@localhost's password:
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-ubuntu-nodemanager-ubuntu.out
```

Fig 1.1: Executing Shell Script

```
-----
-Removing existing directories-
-----
rmr: DEPRECATED: Please use 'rm -r' instead.
18/03/14 10:19:18 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes
Deleted /user/Sqoop/Buys
18/03/14 10:19:19 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes
Deleted /user/Sqoop/Clicks
rmr: DEPRECATED: Please use 'rm -r' instead.
18/03/14 10:19:25 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes
Deleted /user/Pig/Filtered_Buys
18/03/14 10:19:26 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes
Deleted /user/Pig/Filtered_Clicks
```

Fig 1.2: Removing Existing Directories

```
-----
-Importing data from Mysql-
-----
-----
-Importing data from Buys Table-
-----
Warning: /usr/local/sqoop/../hbase does not exist! HBase imports will fail.
Please set $HBASE_HOME to the root of your HBase installation.
Warning: /usr/local/sqoop/../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/local/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /usr/local/sqoop/../zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
18/03/14 10:19:43 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
18/03/14 10:19:44 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
18/03/14 10:19:46 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/03/14 10:19:46 INFO tool.CodeGenTool: Beginning code generation
18/03/14 10:19:51 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `buys` AS t LIMIT 1
18/03/14 10:19:54 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `buys` AS t LIMIT 1
18/03/14 10:19:54 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/local/hadoop
Note: /tmp/sqoop-ubuntu/compile/d0aa1bbb091923ee01be712ae99c7ea7/buys.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
```

Fig 1.3: Importing Data From Buys Table

```
-----
-Importing data from Clicks Table-
-----
Warning: /usr/local/sqoop/../hbase does not exist! HBase imports will fail.
Please set $HBASE_HOME to the root of your HBase installation.
Warning: /usr/local/sqoop/../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/local/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /usr/local/sqoop/../zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
18/03/14 10:21:06 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
18/03/14 10:21:06 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
18/03/14 10:21:08 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/03/14 10:21:09 INFO tool.CodeGenTool: Beginning code generation
18/03/14 10:21:09 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `clicks` AS t LIMIT 1
18/03/14 10:21:09 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `clicks` AS t LIMIT 1
18/03/14 10:21:09 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/local/hadoop
Note: /tmp/sqoop-ubuntu/compile/a927a77fffd5d1e35b62330369488a8c/clicks.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
```

Fig 1.4: Importing Data From Clicks Table

```
-----
----After importing Data----
-----
Found 2 items
-rw-r--r-- 1 ubuntu supergroup          0 2018-03-14 10:20 /user/Sqoop/Buys/_SUCCESS
-rw-r--r-- 1 ubuntu supergroup 27143 2018-03-14 10:20 /user/Sqoop/Buys/part-m-00000
Found 2 items
-rw-r--r-- 1 ubuntu supergroup          0 2018-03-14 10:21 /user/Sqoop/Clicks/_SUCCESS
-rw-r--r-- 1 ubuntu supergroup 22443 2018-03-14 10:21 /user/Sqoop/Clicks/part-m-00000
```

Fig 1.5: After Importing Data

```
-----
-Executing pig scripts-
-----
18/03/14 10:21:59 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/03/14 10:21:59 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/03/14 10:21:59 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2018-03-14 10:21:59,589 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23
2018-03-14 10:21:59,593 [main] INFO org.apache.pig.Main - Logging error messages to: /home/ubuntu/pig_1521003119577.log
2018-03-14 10:22:02,307 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/ubuntu/.pigbootup not fo
2018-03-14 10:22:02,936 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated
e.jobtracker.address
2018-03-14 10:22:02,943 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. I
S
2018-03-14 10:22:02,944 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hado
://localhost:54310
2018-03-14 10:22:04,961 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-pigscript.pig-5eb025f
227314
2018-03-14 10:22:04,962 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set
2018-03-14 10:22:06,836 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. I
S
2018-03-14 10:22:07,803 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: DISTIN
2018-03-14 10:22:08,032 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. I
S
```

Fig 1.6: Executing Pig Scripts

```
-----
----Filtered Data----
-----
Found 2 items
-rw-r--r-- 1 ubuntu supergroup          0 2018-03-14 10:22 /user/Pig/Filtered_Buys/_SUCCESS
-rw-r--r-- 1 ubuntu supergroup 26881 2018-03-14 10:22 /user/Pig/Filtered_Buys/part-r-00000
Found 2 items
-rw-r--r-- 1 ubuntu supergroup          0 2018-03-14 10:23 /user/Pig/Filtered_Clicks/_SUCCESS
-rw-r--r-- 1 ubuntu supergroup 22367 2018-03-14 10:23 /user/Pig/Filtered_Clicks/part-r-00000
```

Fig 1.7: After Filtering Data

```
-----
-Executing hive queries-
-----
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.4.1.jar!/org/slf4j/impl/StaticLoggerBinder.clas
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/Sta
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/usr/local/hive/lib/hive-common-2.1.1.jar!/hive-log4j2.properties As
OK
Time taken: 17.978 seconds
OK
Time taken: 2.779 seconds
Loading data to table default.clicks
OK
Time taken: 2.915 seconds
OK
```

Fig 1.8: Launching Hive Job

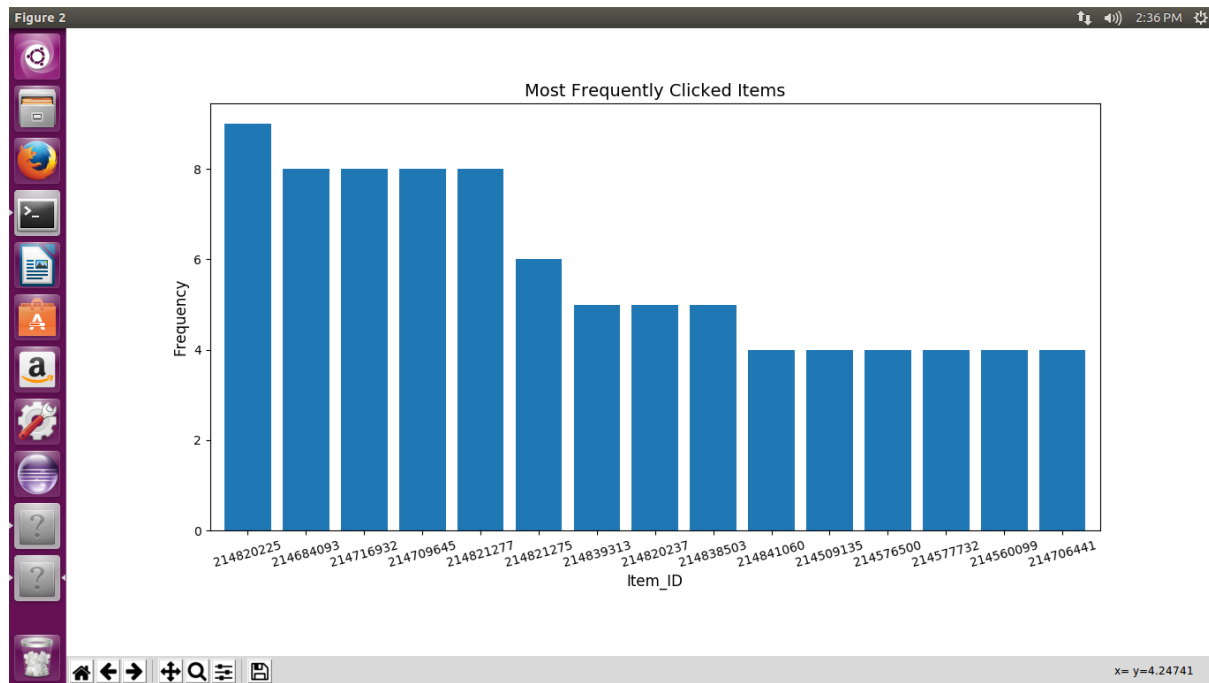


Fig 1.9: Most Frequently Clicked Items

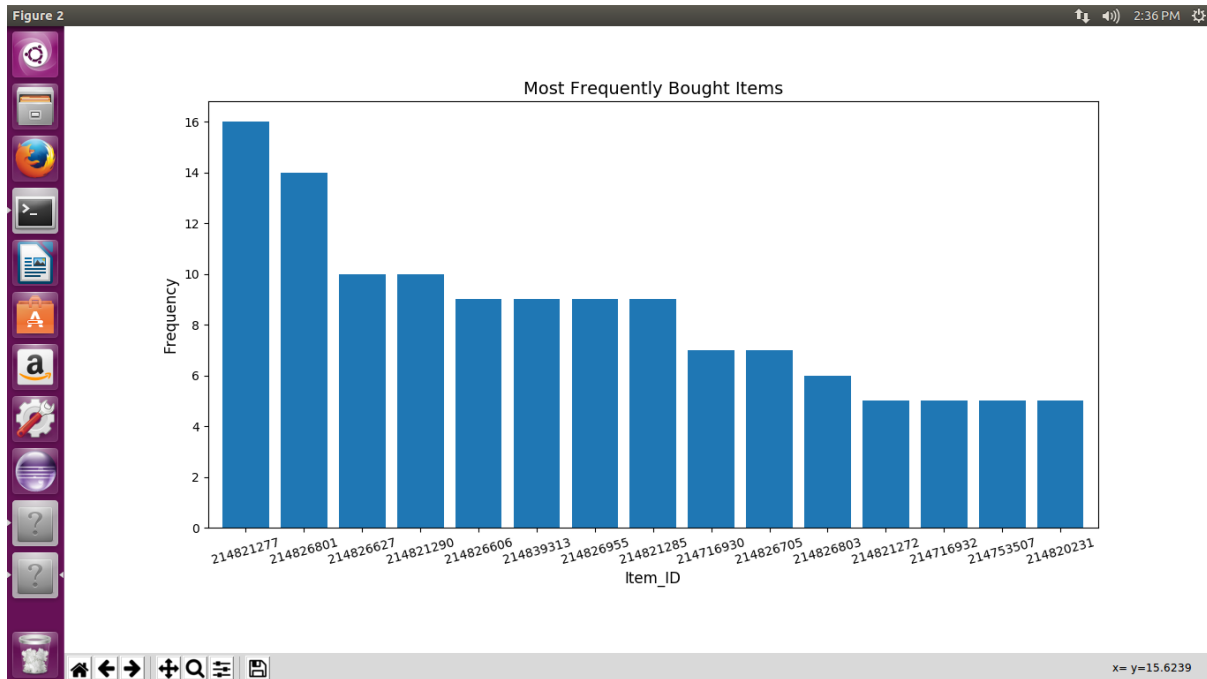


Fig 1.10: Most Frequently Bought Items

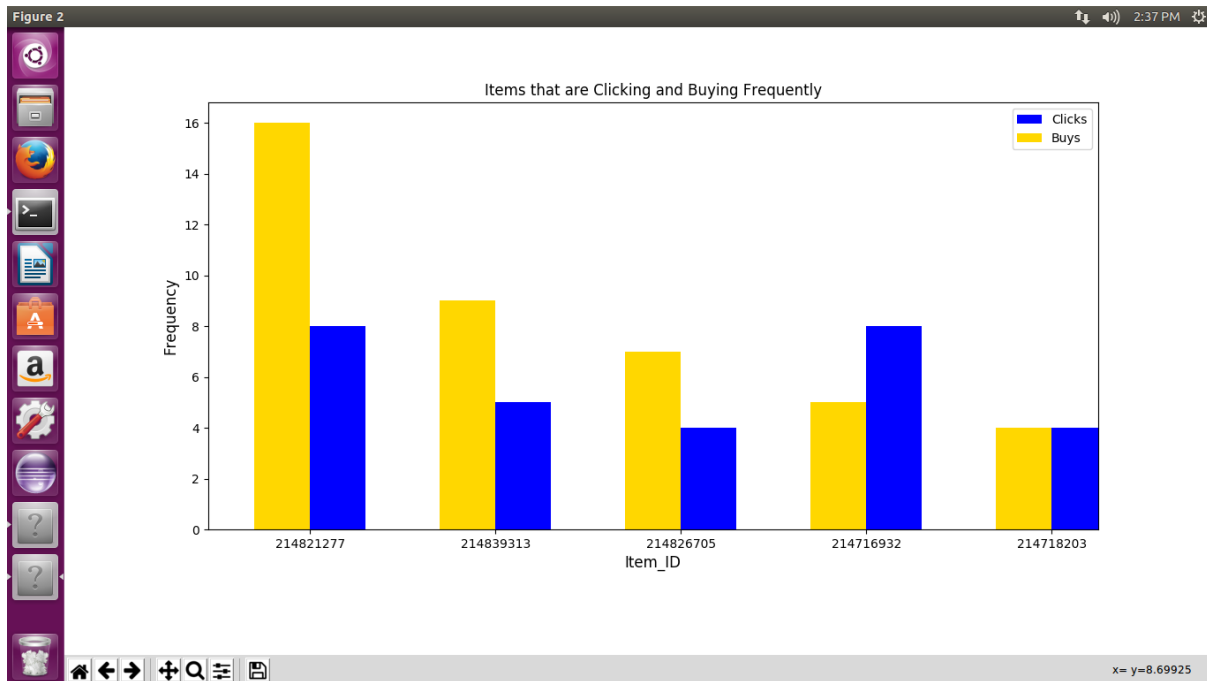


Fig 1.11: Items That are Clicking and Buying Frequently

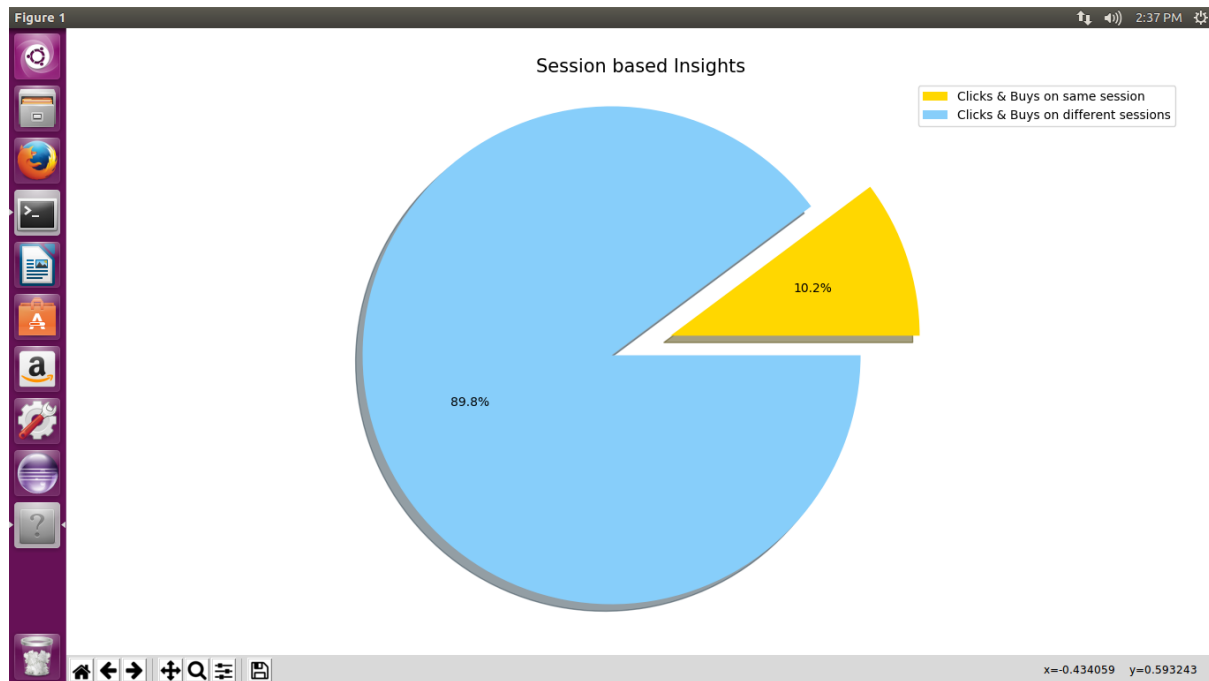


Fig 1.12: Session based Insights

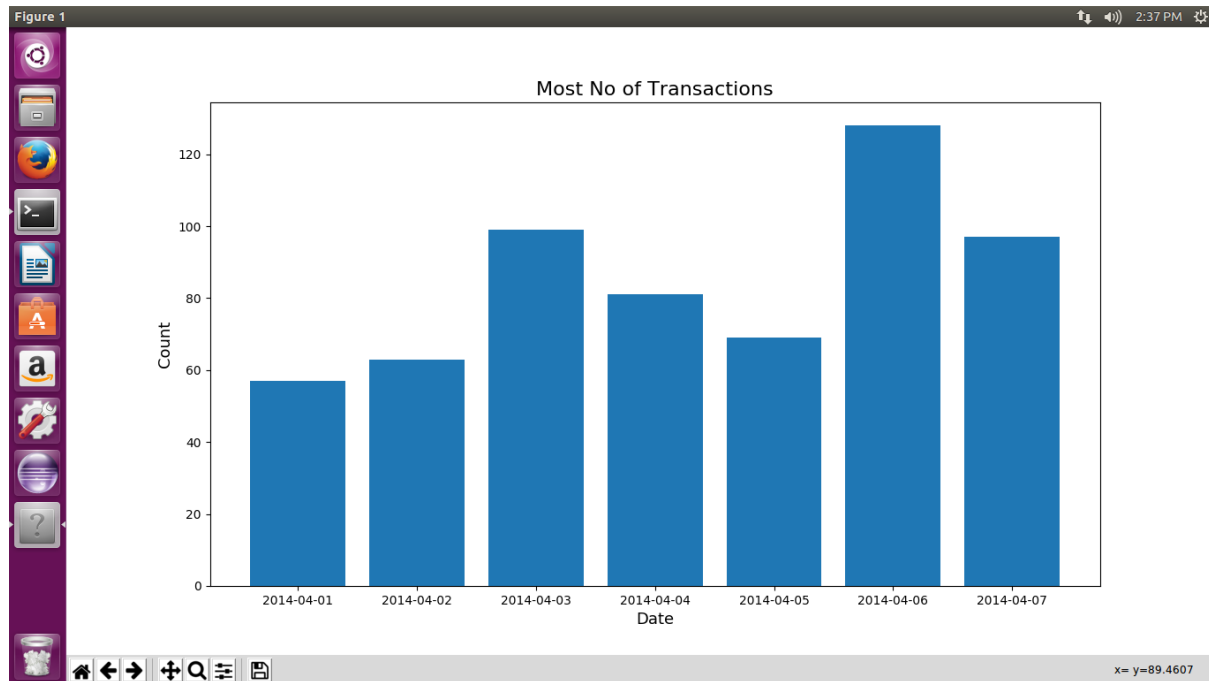


Fig 1.13: Most Number Of Transactions

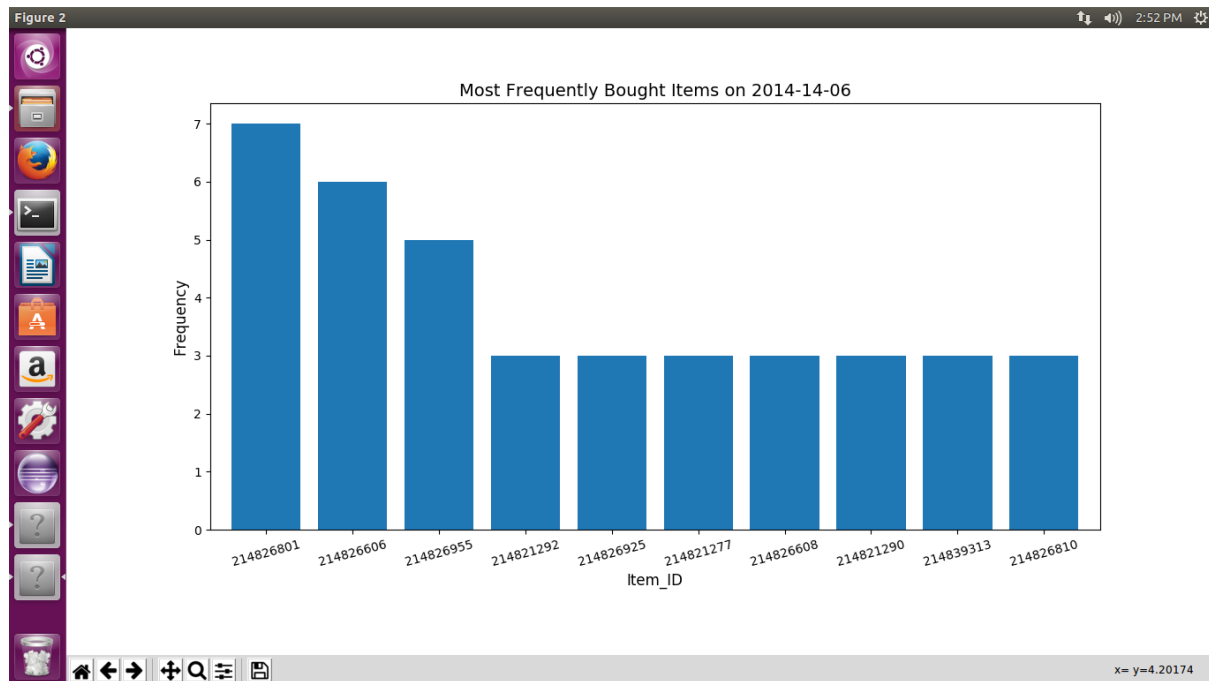


Fig 1.14: Most Frequently Bought Items On 2014-14-06

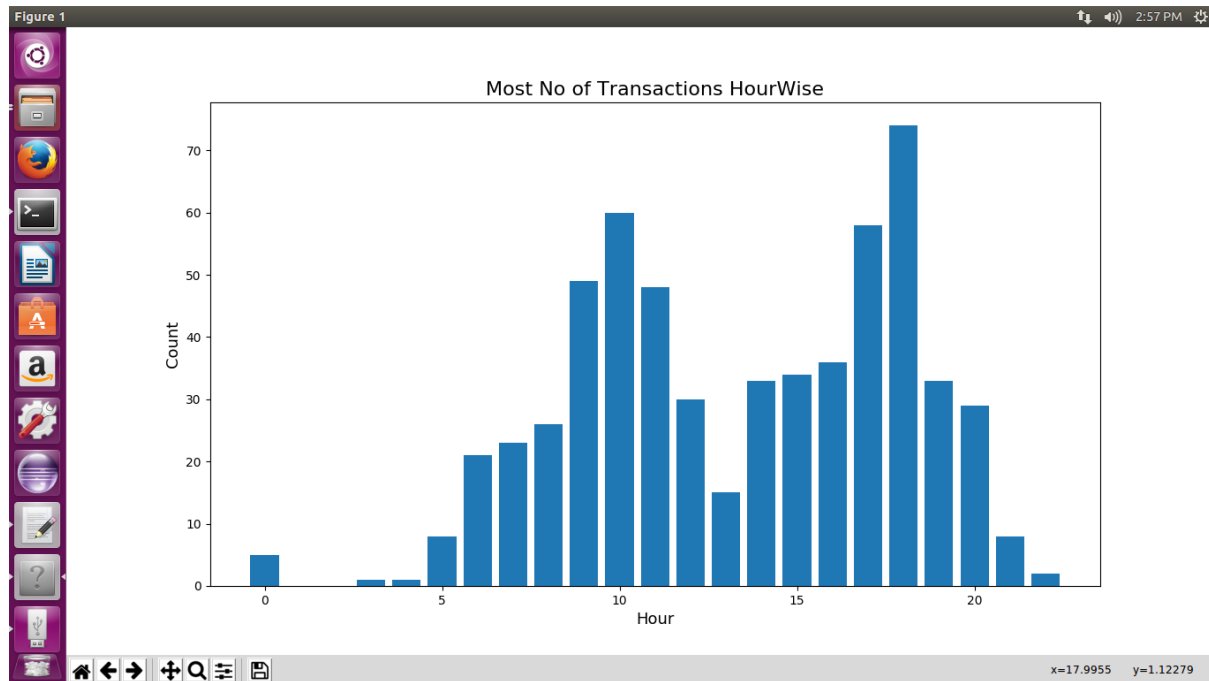


Fig 1.15: Most Number Of Transactions HourWise

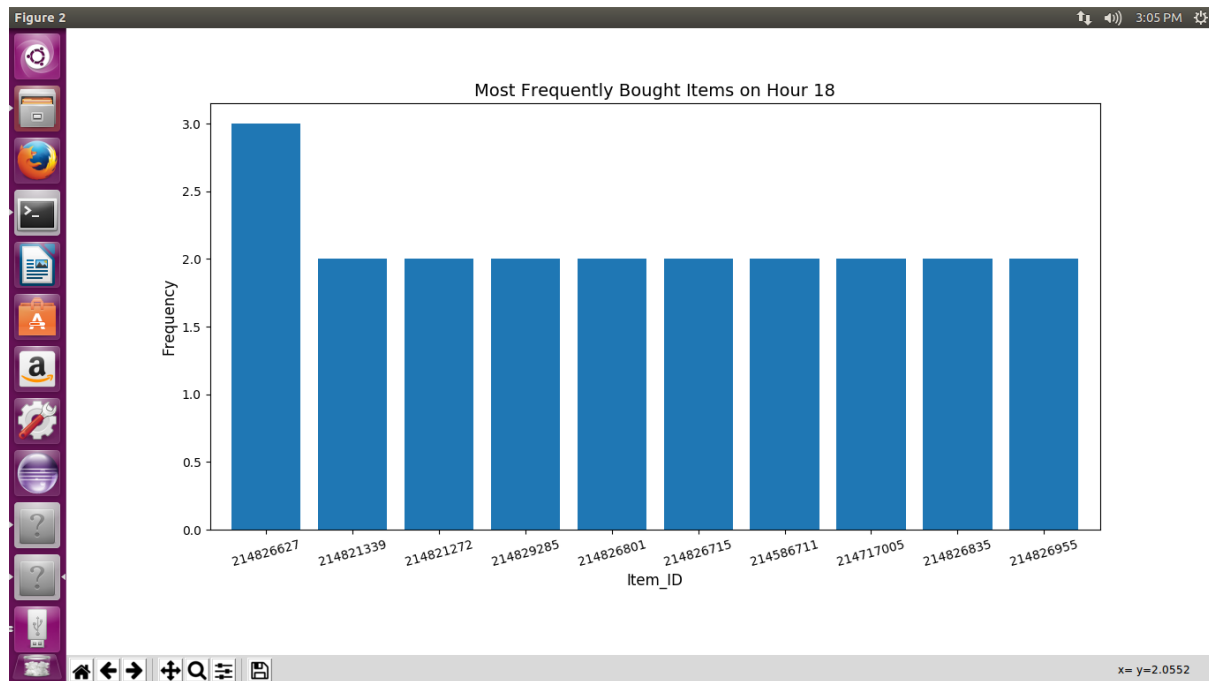


Fig 1.16: Most Frequently Bought Items On Hour 18