Hive table creation

```
hive> CREATE TABLE retail_cleaned (
    > Invoice STRING,
    > StockCode STRING,
    > Description STRING,
    > Quantity INT,
    > InvoiceDate STRING,
    > Price DOUBLE,
    > CustomerID STRING,
    > Country STRING
    > )
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ','
    > STORED AS TEXTFILE;
2025-10-21T14:50:17,010 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 3e271ef7-c8f7-4ac6-b205-40c06d39c1e5
2025-10-21T14:50:17,010 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 3e271ef7-c8f7-4ac6-b205-40c06d39c1e5 main
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. AlreadyExistsException(message:Table hive.default.retail_cleaned already exists)
2025-10-21T14:50:17,419 INFO [3e271ef7-c8f7-4ac6-b205-40c06d39c1e5 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 3e271ef7-c8f7-4ac6-b205-40c06d39c1e5
2025-10-21T14:50:17,420 INFO [3e271ef7-c8f7-4ac6-b205-40c06d39c1e5 main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to  main
hive> SHOW TABLES;
2025-10-21T14:50:59,777 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 3e271ef7-c8f7-4ac6-b205-40c06d39c1e5
2025-10-21T14:50:59,778 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 3e271ef7-c8f7-4ac6-b205-40c06d39c1e5 main
OK
2025-10-21T14:51:00,080 INFO [3e271ef7-c8f7-4ac6-b205-40c06d39c1e5 main] org.apache.hadoop.conf.Configuration.deprecation - mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.input
retail_cleaned
Time taken: 0.293 seconds, Fetched: 1 row(s)
2025-10-21T14:51:00,320 INFO [3e271ef7-c8f7-4ac6-b205-40c06d39c1e5 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 3e271ef7-c8f7-4ac6-b205-40c06d39c1e5
2025-10-21T14:51:00,321 INFO [3e271ef7-c8f7-4ac6-b205-40c06d39c1e5 main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to  main
```

```
hive> LOAD DATA INPATH '/user/zinya/retail_project/cleaned/online_retail_cleaned.csv'
    > INTO TABLE retail_cleaned;
2025-10-21T14:52:57,440 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 3e271ef7-c8f7-4ac6-b205-40c06d39c1e5
2025-10-21T14:52:57,441 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 3e271ef7-c8f7-4ac6-b205-40c06d39c1e5 main
Loading data to table default.retail_cleaned
OK
Time taken: 1.619 seconds
2025-10-21T14:52:59,063 INFO [3e271ef7-c8f7-4ac6-b205-40c06d39c1e5 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 3e271
2025-10-21T14:52:59,063 INFO [3e271ef7-c8f7-4ac6-b205-40c06d39c1e5 main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to  main
hive> SELECT * FROM retail_cleaned LIMIT 10;
2025-10-21T14:53:38,651 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 3e271ef7-c8f7-4ac6-b205-40c06d39c1e5
2025-10-21T14:53:38,651 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 3e271ef7-c8f7-4ac6-b205-40c06d39c1e5 main
2025-10-21T14:53:41,619 INFO [3e271ef7-c8f7-4ac6-b205-40c06d39c1e5 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://l
b205-40c06d39c1e5/hive_2025-10-21_14-53-38_675_9118229916584099941-1/-mr-10001/.hive-staging_hive_2025-10-21_14-53-38_675_9118229916584099941-1
2025-10-21T14:53:41,951 INFO [3e271ef7-c8f7-4ac6-b205-40c06d39c1e5 main] org.apache.hadoop.conf.Configuration.deprecation - mapred.task.is.map is deprecated. Inste
OK
"Invoice","StockCode","Description","Quantity","InvoiceDate","Price","Customer.ID","Country"
"489434","85048","15CM CHRISTMAS GLASS BALL 20 LIGHTS",12,"2009-12-01 07:45:00",6.95,13085,"United Kingdom"
"489434","79323P","PINK CHERRY LIGHTS",12,"2009-12-01 07:45:00",6.75,13085,"United Kingdom"
"489434","79323W"," WHITE CHERRY LIGHTS",12,"2009-12-01 07:45:00",6.75,13085,"United Kingdom"
"489434","22041","RECORD FRAME 7"" SINGLE SIZE ",48,"2009-12-01 07:45:00",2.1,13085,"United Kingdom"
"489434","21232","STRAWBERRY CERAMIC TRINKET BOX",24,"2009-12-01 07:45:00",1.25,13085,"United Kingdom"
"489434","22064","PINK DOUGHNUT TRINKET POT ",24,"2009-12-01 07:45:00",1.65,13085,"United Kingdom"
"489434","21871","SAVE THE PLANET MUG",24,"2009-12-01 07:45:00",1.25,13085,"United Kingdom"
"489434","21523","FANCY FONT HOME SWEET HOME DOORMAT",10,"2009-12-01 07:45:00",5.95,13085,"United Kingdom"
"489435","22350","CAT BOWL ",12,"2009-12-01 07:46:00",2.55,13085,"United Kingdom"
Time taken: 3.3 seconds, Fetched: 10 row(s)
2025-10-21T14:53:42,062 INFO [3e271ef7-c8f7-4ac6-b205-40c06d39c1e5 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 3e2
2025-10-21T14:53:42,063 INFO [3e271ef7-c8f7-4ac6-b205-40c06d39c1e5 main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to  main
hive>
```

Hive udf

```
hive> CREATE TEMPORARY MACRO classify_demand(quantity BIGINT)
    > CASE
    >     WHEN quantity IS NULL THEN 'Unknown'
    >     WHEN quantity >= 10000 THEN 'High'
    >     WHEN quantity >= 2000 THEN 'Medium'
    >     ELSE 'Low'
    > END;
2025-10-25T01:17:07,991 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 718bd7c4-40ad-4c82-bd2b-72a97bcc7319
2025-10-25T01:17:08,007 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 718bd7c4-40ad-4c82-bd2b-72a97bcc7319 main
OK
Time taken: 0.574 seconds
2025-10-25T01:17:08,612 INFO [718bd7c4-40ad-4c82-bd2b-72a97bcc7319 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 718bd7c4-40ad-4c82-bd2b-72a97bcc7319
2025-10-25T01:17:08,612 INFO [718bd7c4-40ad-4c82-bd2b-72a97bcc7319 main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to  main
hive>
```

```
product_demand_category.stockcode    product_demand_category.total_quantity  product_demand_category.demand_category
84077    110249  High
85123A   100152  High
85099B   98356   High
21212    96560   High
22197    87203   High
84879    81817   High
23166    78033   High
17003    71440   High
21977    56794   High
84991    54716   High
22492    45665   High
15036    44485   High
20725    40942   High
84755    40839   High
22178    40748   High
21213    40733   High
22616    40367   High
22386    39840   High
21232    38983   High
84568    38574   High
Time taken: 1.891 seconds, Fetched: 20 row(s)
2025-10-25T01:23:04,908 INFO [718bd7c4-40ad-4c82-bd2b-72a97bcc7319 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 718bd7c4-40ad-4c82-bd2b-72a97bcc7319
2025-10-25T01:23:04,908 INFO [718bd7c4-40ad-4c82-bd2b-72a97bcc7319 main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to  main
hive>
```

Udf output

```
85099F  JUMBO BAG STRAWBERRY      37127
22197   POPCORN HOLDER  36761
22086   PAPER CHAIN KIT 50'S CHRISTMAS  36581
84992   72 SWEETHEART FAIRY CAKE CASES  35438
21982   PACK OF 12 SUKI TISSUES       32827
84347   ROTATING SILVER ANGELS T-LIGHT HLDR    32525
21975   PACK OF 60 DINOSAUR CAKE CASES 31575
85099C  JUMBO  BAG BAROQUE BLACK WHITE 31249
23084   RABBIT NIGHT LIGHT        30788
71459   HANGING JAM JAR T-LIGHT HOLDER  30764
16014   SMALL CHINESE STYLE SCISSOR    30370
21731   RED TOADSTOOL LED NIGHT LIGHT   29715
21931   JUMBO STORAGE BAG SUKI  29505
22469   HEART OF WICKER SMALL   29079
22952   60 CAKE CASES VINTAGE CHRISTMAS 28744
47566   PARTY BUNTING   28395
22355   CHARLOTTE BAG SUKI DESIGN       28290
21790   VINTAGE SNAP CARDS      27942
22151   PLACE SETTING WHITE HEART       27897
22411   JUMBO SHOPPER VINTAGE RED PAISLEY      27733
22423   REGENCY CAKESTAND 3 TIER       27594
20727   LUNCH BAG  BLACK SKULL. 27126
84970S  HANGING HEART ZINC T-LIGHT HOLDER      26844
21981   PACK OF 12 WOODLAND TISSUES    26472
21929   JUMBO BAG PINK VINTAGE PAISLEY 25954
21985   PACK OF 12 HEARTS DESIGN TISSUES       25708
37410   BLACK AND WHITE PAISLEY FLOWER MUG     25707
22386   JUMBO BAG PINK POLKADOT 25627
21984   PACK OF 12 PINK PAISLEY TISSUES        25509
20719   WOODLAND CHARLOTTE BAG  24948
Time taken: 104.286 seconds, Fetched: 50 row(s)
2025-10-21T19:34:34,639 INFO [a8d98a99-877c-4cc8-ba37-b75dd61ab27e main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: a8d98a99-877c-4cc8-ba37-b75dd61ab27e
2025-10-21T19:34:34,640 INFO [a8d98a99-877c-4cc8-ba37-b75dd61ab27e main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to  main
hive>
```

Hive table creation

```
hive> SELECT * FROM retail_cleaned LIMIT 5;
2025-10-21T15:59:12,750 INFO [827eadb4-be44-41b6-a5e6-51ba94d36083 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 827eadb4-be44-41b6-a5e6-51ba94d36083
2025-10-21T15:59:13,766 WARN [827eadb4-be44-41b6-a5e6-51ba94d36083 main] org.apache.hadoop.hive.ql.session.SessionState - METASTORE_FILTER_HOOK will be ignored, since hive.security.authorization.manager is set t
) instance of HiveAuthorizerFactory.
2025-10-21T15:59:17,138 INFO [827eadb4-be44-41b6-a5e6-51ba94d36083 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/tmp/hive/zinya/827eadb4-be44-41b6-
5e6-51ba94d36083/hive_2025-10-21_15-59-12_944_6942786901707660495-1/-mr-10001/.hive-staging_hive_2025-10-21_15-59-12_944_6942786901707660495-1
2025-10-21T15:59:17,529 INFO [827eadb4-be44-41b6-a5e6-51ba94d36083 main] org.apache.hadoop.conf.Configuration.deprecation - mapred.task.is.map is deprecated. Instead, use mapreduce.task.ismap
OK
2025-10-21T15:59:17,596 INFO [827eadb4-be44-41b6-a5e6-51ba94d36083 main] org.apache.hadoop.conf.Configuration.deprecation - mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.inputdir
nvoice StockCode       Description     NULL    InvoiceDate     NULL    CustomerID      Country
89434  85048   15CM CHRISTMAS GLASS BALL 20 LIGHTS     12      01/12/2009 07:45        6.95    13085   United Kingdom
89434  79323P  PINK CHERRY LIGHTS      12      01/12/2009 07:45        6.75    13085   United Kingdom
89434  79323W  WHITE CHERRY LIGHTS     12      01/12/2009 07:45        6.75    13085   United Kingdom
89434  22041   "RECORD FRAME 7"" SINGLE SIZE " 48      01/12/2009 07:45        2.1     13085   United Kingdom
Time taken: 4.722 seconds, Fetched: 5 row(s)
2025-10-21T15:59:18,161 INFO [827eadb4-be44-41b6-a5e6-51ba94d36083 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 827eadb4-be44-41b6-a5e6-51ba94d36083
```

How product demand categorization guides inventory and supply chain strategy

Product demand categorisation into High, Medium, and Low segments provides actionable intelligence for optimising inventory management and supply chain operations. This data-driven approach enables retailers to allocate resources efficiently and reduce costs while maintaining customer satisfaction. High demand products are the revenue drivers, and they require priority treatment. Inventory strategy should focus on maintaining stick levels to prevent stockouts, which directly impact revenue and customer loyalty. Supply chain implications include establishing direct supplier relationships, negotiating bulk purchase prioritise accessibility and these items warrant expedited shipping arrangements to ensure constant availability. The medium demand products are steady performers require balanced management. Inventory approach involves regular monitoring with moderate stock levels based on predictable demand patterns. Supply chain strategy should implement scheduled reordering cycles and maintain standard supplier relationships. These products benefit from demand forecasting models to optimise order quantities and reduce holding costs while avoiding stockouts. The low demand products are slow-moving items tie up capital and warehouse space. Inventory decisions should focus on minimising stock levels, implementing just-in-time ordering, or considering discontinuation. Supply chain optimization involves consolidating orders with other products to reduce shipping costs

and negotiating extended lead times with suppliers. Marketing teams should evaluate promotional strategies to boost sales or plan clearance activities to free up resources. This categorisation enables data-driven decisions across procurement, warehousing, and logistics. It reduces inventory holding costs by right-sizing stock levels, improves cash flow by preventing over-investment in slow movers, and enhances customer satisfaction by ensuring high-demand products remain available. The result is a more responsive, efficient supply chain aligned with actual market demand.