Bikal Basnet

Basnet.beekal@gmail.com

Abstract

You can **increase your Sales by upto 10 -35%** as the Amazon did. With right set of recommendation tools and suggestions, it is now possible.

Product Recommendation System  
 for XXX Using Big Data Technology   
(Hadoop, Mahout, SSIS and Pig)   
v 1.0

Proof of Concept

Contents

[Introduction : 2](#_Toc408938139)

[Why Recommendation System : 2](#_Toc408938140)

[Why Mahout ? 3](#_Toc408938141)

[Limitation : 3](#_Toc408938142)

[Data size : 4](#_Toc408938143)

[Results: 4](#_Toc408938144)

[References 10](#_Toc408938145)

[Appendix : 11](#_Toc408938146)

[Full Recommendation Result Set 11](#_Toc408938147)

[Code 11](#_Toc408938148)

# Introduction :

This report contains product recommendation for individual skygeek customers. The product recommendation is performed taking into consideration the following parameters

* The product that a customer purchases
* Similar customer tend to buy similar products i.e Group similar users together and based upon the group sharing similar product purchase history, recommend the products.

The recommendation system was implemented using Mahout, a scalable machine learning library for the Big Data technology.

# Why Recommendation System :

“Recommendation system to a company, is what a blood is to a Human Being”

The statement can simply be justified by looking into the fact that, Amazon increased their sales by 29% to $12.83 billion in its second fiscal quarter up from $9.9 billion during the same time last year, with its use of the recommendation system(Fortune.com)（Ama14）.

Not only that numerous experiences and numerous research papers have pointed out that, sales are expected to increase in between 10 to 35 percent. According to 2006 sales figure, 35 % of the Amazons sales were done through their recommendation system. （Rec14）

The importance of the recommender system can also be gauged by the fact that “NetFlix ran a

“One Million Dollar” Cash Prize Competition to with sole focus to improve its recommendation system only.

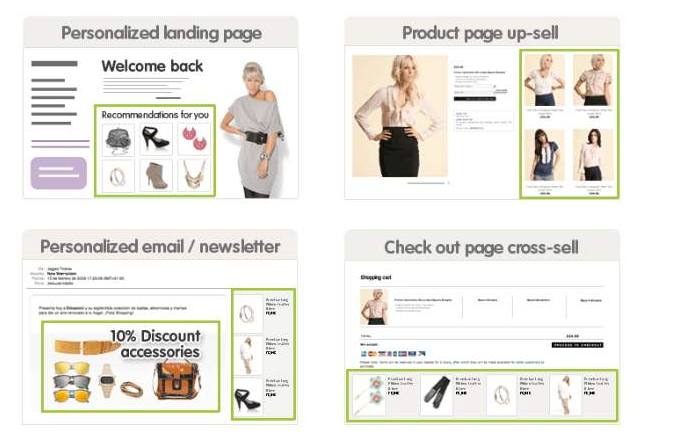


Fig. Showing Recommendation System usage at a Mock clothing store.

They not only increase sales, but also encourage customers to return to business for future purchases.

# Why Mahout ?

While it was possible to implement the recommendation system using R, Statistical programming language for small data size. Scaling up for more than 350,000 customers across more than 150,000 product items presented significant challenge and could not be addressed with R.

Taking into consideration our data size and the scalability desired, the Mahout was choosen.

# Limitation :

Recommendation for only the top 25,000 customers has been performed.

To perform recommendation across all the user sets, higher computation power is required.

Avialiability of Amazon Web Servers would enable recommendation for all customers and would add significantly to the Companies Skill Set list and as a skill set showroom for the valued customers.

# Data size :

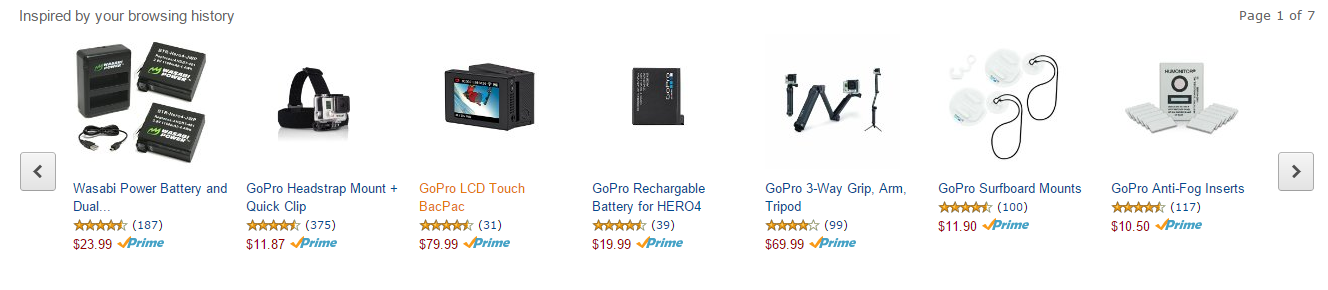
For the recommendation system, all the customers data and their product purchase history till date were analysed. More than 350,000 Customers (Customer Table of SEOM) and their purchase history (Order Table and Order Details Table) was taken into account as the input data for the recommender algorithm.

# Results:

Below mentioned is a sample of top 10 recommendations for 10 customers.

Presenting top 10 recommendations for all 25000 customers, since is not feasible in the report, separate file containing the recommendation set is added to the appendix section of the report.

**Recommendations for you**



(Recommendations as in the sample picture shown above)

|  |  |  |
| --- | --- | --- |
| CustomerId | Product Name | Recommendation  Priority or Purchase Score |
| 250001 | FlyBoys FB1000 Flexi-Lock Checklist Ring ( 10 Pack ) | 4 |
| 250001 | FlyBoys FB1000 Opaque Flexi-Lock Checklist Ring, 10 Pack | 4 |
| 250001 | FlyBoys FB1000 Red Flexi-Lock Checklist Ring, 10 Pack | 4 |
| 250001 | FlyBoys FB1000 Flexi-Lock Checklist Ring, Black | 4 |
| 250001 | CarProp License Plate Propeller-BLUE | 4 |
| 250001 | Cessna Aircraft Montage Christmas Tree Ornament | 4 |
| 250001 | ATP Alaska Terminal Procedure - Bound - Expires 09/18/2014 | 4 |
| 250001 | SkyGeek PP-9 Wooden 9 Paint Mixing Paddle / Stick - Quart - Each" | 4 |
| 250001 | MilSpec MS-F4C #4 Flush Head Z-Spec Grommet Flaring Tool | 4 |
| 250001 | MilSpec MS-ZC C-Clamp for All Z-Spec Grommet Flaring Tool Sets | 4 |
| 250027 | AFDEC East Central U.S. Airport/Facility Directory (AFD) - Expires 09/18/2014 | 13 |
| 250027 | SCHA Charlotte Sectional Chart - NEXT EDITION | 13 |
| 250027 | TPPSE4 Approach Chart Southeast Vol. 4 - Loose Leaf - Expires 09/18/2014 | 12.29566 |
| 250027 | ELUS33 L-33/34 IFR Low Altitude Enroute Chart - Expires 09/18/2014 | 12.19516 |
| 250027 | SBIL Billings Sectional Chart - Expires 08/21/2014 | 11.89488 |
| 250027 | SMEM Memphis Sectional Chart - Expires 09/18/2014 | 11.8847 |
| 250027 | TATL Atlanta Terminal Area Chart - Expires 08/21/2014 | 11.6553 |
| 250027 | SLH Lake Huron Sectional Chart - Expires 09/18/2014 | 11.45148 |
| 250027 | AFDSC South Central U.S. Airport/Facility Directory (AFD) - Expires 09/18/2014 | 11.43359 |
| 250027 | ELUS29 L-29/30 IFR Low Altitude Enroute Chart - Expires 09/18/2014 | 11.3688 |
| 250031 | AVBLEND Aircraft Engine Oil Additive - 12 Oz. Can | 1 |
| 250031 | SWAS Washington Sectional Chart - Expires 02/05/2015 | 1 |
| 250031 | Champion Slick 11-8150-1 Magneto Timing Tool | 1 |
| 250031 | Aircraft Tool Supply 4-Piece Bucking Bar Set | 1 |
| 250031 | Aeroshell 15W-50 Multigrade Aircraft Engine Oil - Quart Bottle | 1 |
| 250031 | K.P. Products FS01 Rivet Fan Spacing Tool | 1 |
| 250031 | TNY New York Terminal Area Chart - Expires 11/13/2014 | 1 |
| 250031 | Phillips XC 20W50 Aviation Oil - Quart Bottle | 1 |
| 250031 | Rapco 212CW Dry Air Pump Overhaul - Core Exchange Fee +$100 | 1 |
| 250031 | Aircraft Tool Supply ATS450-WG Repalacement Wire Pack for ATS450 Spark Plug Feeler Gauge | 1 |
| 250040 | Champion Slick K3822 Slick T15 Assembly | 18.97413 |
| 250040 | Military Standard MS20470AD6-16 Rivet Solid - Univ Hd - Al - 1Lb | 18.72706 |
| 250040 | Piper 63310-00 Spacer | 18.72706 |
| 250040 | Fitting, STR Female Flare, Aluminum | 18.50718 |
| 250040 | National Aerospace Standard NAS1329A06K75 Rivnut Flt Hd - Al | 18.31281 |
| 250040 | Parker Hannifin Airborne 2H3-16 Valve Air Regulator | 18.31281 |
| 250040 | Military Standard MS29513-230 O-Ring | 18.31281 |
| 250040 | Teledyne Continental 10-391586 Ignition Block | 18.31281 |
| 250040 | TCO Manufacturing ES4C Static Wick | 18.22145 |
| 250040 | Cleveland 067-03400 Axle Spacer | 18.22145 |
| 250043 | Cummins 4089728 Fuel Injector | 3.378444 |
| 250043 | Military Standard MS25083-7BB12 Electrical Lead - Aluminum | 3.378444 |
| 250043 | Parker Cleveland 155-14200 Ring Retainer | 3.378444 |
| 250043 | Powerline 36 Insulated Wand - JE5170D36" | 3.378444 |
| 250043 | Clamp | 3.378444 |
| 250043 | Slide Assembly | 3.378444 |
| 250043 | 7/8 Torque Adapter" | 3.378444 |
| 250043 | National Aerospace Standard NAS1133E6P Screw, Machine | 3.378444 |
| 250043 | Cutoff Wheel Holder | 3.378444 |
| 250043 | Navigation Light, Wing Tip | 3.378444 |
| 250052 | Phillips 66 X/C 25W-60 Radial Engine Aircraft Oil - Gallon Jug | 1 |
| 250070 | Nut, Hex, Self Locking | 1 |
| 250070 | Hose Assembly, Oil - AE710531-1 | 1 |
| 250070 | Bolt, Hex, Close Tolerance, Crescent - AN174C12A | 1 |
| 250070 | Nut | 1 |
| 250070 | Nut | 1 |
| 250070 | Gasket - Pack of 6 | 1 |
| 250070 | MSDS Hazmat Certification Required with Shipment | 1 |
| 250070 | Arrow-Magnolia CD-2035 Carbon-X Aircraft Soap - 1 Gallon Jug | 1 |
| 250070 | AeroShell Grease 22 Advanced General Purpose Aircraft Grease - 14 Oz. Cartridge - MIL-PRF-81322G | 1 |
| 250088 | ICOM BP-208N Alkaline Battery for use with ICOM IC-A6 & IC-A24 | 1 |
| 250088 | AFDSE South East U.S. Airport/Facility Directory (AFD) - Expires 09/18/2014 | 1 |
| 250088 | Pilot USA PA-97 General Aviation Headset Plug to Airbus Headset Plug Adapter | 1 |
| 250088 | SATL Atlanta Sectional Chart - Expires 08/21/2014 | 1 |
| 250088 | SkyGeek Same Day Ship Guarantee | 1 |
| 250088 | Avcomm P3-P01 Non-Padded Headset Bag | 1 |
| 250088 | SJAC Jacksonville Sectional Chart - Expires 08/21/2014 | 1 |
| 250088 | Vertex VXA-300 Pilot III NAV-COM Transceiver 110 VAC | 1 |
| 250088 | WCH25 CH-25 World Aeronautical Charts - Expires 12/11/2014 | 1 |
| 250088 | FuelHawk C-152/12 Cessna C152 12 Gallon Tank Fuel Stick | 1 |
| 250099 | SkyGeek.Com Small T-Shirt | 1 |
| 250099 | 3M™ Aerospace Sealant AC-350 Class B-1/2 Quick Cure Fuel Tank & Fuselage - 2.5 oz Kit | 1 |
| 250099 | Jet Stream Aviation PREP01 PBS De-Ice Boot Prep - 1 Quart Bottle | 1 |
| 250099 | Goodrich 74-451-179 ShineMaster Prep Cleaner - Quart | 1 |
| 250099 | Aeroshell W65 SAE 30 Aircraft Oil -12 Quart Case | 1 |
| 250099 | AvLabs AVL-CV-16 Clear View Plastic Polish & Protectant - 13 oz Aerosol Can | 1 |
| 250099 | SkyGeek.com AM-KIT Aircraft Cleaning Kit (SALE-PRICE) | 1 |
| 250099 | B & B Tritech 3100 Compressor Wash - 5 Gallon Compressor Wash - MIL-C-85704, Type I | 1 |
| 250099 | Arrow Magnolia CE-5901 32 oz. Trigger Spray Bottle | 1 |
| 250099 | AeroShell Fluid 3 General Purpose Mineral Lubricating Oil - Gallon - MIL-PRF-7870D | 1 |
| 250106 | ASA Red Softcover Engine Log Book | 2.235333 |
| 250106 | Alaska's Bush Pilots The Real Deal DVD (CLEARANCE) | 2.228178 |
| 250106 | Grez-Off 22732 Heavy Duty Degreaser - 32 oz. Trigger Spray | 2.228178 |
| 250106 | 3M Peltor 97023-00000 Blue Junior Ear Muffs | 2.228178 |
| 250106 | Hot Wings 14121 SR-71 Black Bird Red Striped Diecast Aircraft | 2.228178 |
| 250106 | Met All TC-10 Aluminum Polish - 16 Oz. Can - MIL-P-6888C Type II | 2.228178 |
| 250106 | ASA 2014 FAR/AIM Regulations for General Aviation, Sport Pilots, & Instructors | 2.228178 |
| 250106 | Hot Wings 14139 MIG-29 Russian Air Force Diecast Aircraft Toy/Collectible | 2.228178 |
| 250106 | Designer Decal 5066 I Love Airplane Noise" Bumper Sticker" | 2.228178 |
| 250106 | DURACELL AA Size Alkaline Battery MN1500, 4 Pack | 2.228178 |
| 250147 | DURACELL AA Size Alkaline Battery MN1500, 4 Pack | 1 |
| 250147 | Aeroshell W100 Plus SAE 50 Aircraft Oil - Quart Bottle | 1 |
| 250147 | Sensata Klixon 7271-8-3 Circuit Breaker - 3 Amp | 1 |
| 250147 | Skurka Aerospace 150SG1009-20XL2 Brush | 1 |
| 250147 | Gill 3 Quart Pack Aircraft Battery Electrolyte Fluid | 1 |
| 250147 | LITHIUM BATTERY PACK FOR ELT 2 | 1 |
| 250147 | Sky-Tec High-Torque 149-24HT Lycoming Starter - 24 Volt - Standard - Lyc PN 31B22111 | 1 |
| 250147 | Phillips XC 20W50 Aviation Oil - Quart Bottle | 1 |
| 250147 | Foggles Yellow IFR Training Glasses | 1 |
| 250147 | General Electric 4553 Sealed Beam Aircraft Lamp | 1 |
| 250169 | Gill 3 Quart Pack Aircraft Battery Electrolyte Fluid | 2.963678 |
| 250169 | Aeroshell W100 Plus SAE 50 Aircraft Oil - 12 Quart Case | 2.730244 |
| 250169 | Aeroshell 15W-50 Multigrade Aircraft Engine Oil - 5 Liter | 2.578432 |
| 250169 | Hot Wings 17109 B-24 Liberator Diecast Aircraft | 2.370597 |
| 250169 | TATL Atlanta Terminal Area Chart - Expires 08/21/2014 | 2 |
| 250169 | ELUSSET Entire United States L1/2 thru L35/35 IFR Low Altitude Enroute Chart Set - NEXT EDITION | 2 |
| 250169 | EHUSSET Lower 48 United States High Altitude Enroute Chart Set - NEXT EDITION | 2 |
| 250169 | ELUS17 L-17/18 IFR Low Altitude Enroute Chart - NEXT EDITION | 2 |
| 250169 | TTAM Tampa/Orlando Terminal Area Chart - Expires 08/21/2014 | 2 |
| 250169 | Aviation Supplies & Academics ASA-SA-1 Green Soft Cover Aircraft Logbook | 1.810226 |
| 250188 | Guillow's 60 Strato Streak Plane Balsa | 7.469461 |
| 250188 | Guillow's 75 Flying Machine Balsa Airplane | 7.451219 |
| 250188 | HELNY New York Helicopter Route Chart - Expires 04/28/2016 | 7.426433 |
| 250188 | Aeroshell 15W-50 Multigrade Aircraft Engine Oil - Quart Bottle | 7.414541 |
| 250188 | SMON Montreal Sectional Chart - Expires 08/21/2014 | 7.413475 |
| 250188 | GATS Jar Fuel Sampler - 12 oz Capacity | 7.411594 |
| 250188 | ASA 5/8 Remove Before Flight Pitot Tube Cap" | 7.410993 |
| 250188 | WCF19 CF-19 World Aeronautical Charts - Expires 04/02/2015 | 7.405222 |
| 250188 | Kimberly-Clark 34705 KIMTECH 2-Ply Delicate Task Wipers - 119 Count Box | 7.403574 |
| 250188 | AVBLEND Aircraft Engine Oil Additive - 12 Oz. Can | 7.401868 |
| 250196 | TCIN Cincinnati Terminal Area Chart - NEXT EDITION | 3 |
| 250196 | AFDNE NorthEast U.S. Airport/Facility Directory (AFD) - Expires 09/18/2014 | 3 |
| 250196 | SATL Atlanta Sectional Chart - Expires 08/21/2014 | 3 |
| 250196 | TCIN Cincinnati Terminal Area Chart - Expires 12/11/2014 | 3 |
| 250196 | SNY New York Sectional Chart - Expires 11/13/2014 | 3 |
| 250196 | TWAS Baltimore/Washington Terminal Area Chart - Expires 02/05/2015 | 3 |
| 250196 | SCHA Charlotte Sectional Chart - Expires 01/08/2015 | 3 |
| 250196 | SCHI Sectional Chart, Chicago - Expires 10/16/2014 | 3 |
| 250196 | TBOS Boston Terminal Area Chart - Expires 11/13/2014 | 3 |
| 250196 | BTPPNE3 North East Vol. 3 Approach Plates - Bound - Expires 09/18/2014 | 3 |
| 250200 | Pan American Tool 01-030 Cobalt Jobbers Length Drill Bit - Split Point - #30 | 2.204563 |
| 250200 | Pan American Tool 01-040 Cobalt Jobbers Length Drill Bit - Split Point - #40 | 2.204563 |
| 250200 | Lacing Twine | 1 |
| 250200 | Adhesive Tape | 1 |
| 250200 | Skilcraft 8125-00-488-7952 Recyclable Plastic Trigger Spray Bottle - 16 fl oz Capacity | 1 |
| 250200 | Flexible Metal Tube Assy | 1 |
| 250200 | 3M 051131-06481 Black MP3526N/MP3527N Hook and Loop Fastener - 1 x 5 Yard Box" | 1 |
| 250200 | Dummy Load, Electrical | 1 |
| 250200 | Mechatronics Blower Fan | 1 |

# References

Amazon's REcommendation Secret．[Online]http://fortune.com/2012/07/30/amazons-recommendation-secret/．

Movie Recommender System for Profit Maximisatio, Research Paper．[Online]http://azariaa.com/Content/Publications/Movie.pdf．

Recommendation Systems: Increasing Profit by Long Tail．[Online]http://en.webrazzi.com/2009/09/18/recommendation-systems-increasing-profit-by-long-tail/．

# Appendix :

## Full Recommendation Result Set

The full recommendation result set is too big to be attached to the document and hence separate file has been appended to it.

## Code

Data Preparation

create table SeomRecomSys (

      CustomerId int ,

      Product varchar(max),

      ProductsPerCustomer int

)

INSERT into SeomRecomSys (

 CustomerId ,

 Product ,

 ProductsPerCustomer

)

SELECTcust.CustomerID , inv. PK , count( SKU) as products\_per\_customer

 FROM [SEOM].[dbo] .[Orders] ord inner join Customers cust on ord.CustomerID = cust.CustomerID

 inner join [Order Details] orddet on ord. OrderNumber = orddet.OrderNumber

 INNER JOIN Inventory inv on orddet.SKU = inv.LocalSKU

 WHERE orddet.Adjustment = 0 and orddet .UID is not null

 and lower (orddet .product ) != 'packing' and lower(orddet .product ) != 'manufacturer certification'

 and orddet. SKU !='Shipping'

 and orddet. SKU !='Discount'

 and orddet. SKU !='mfg-cert'

 and orddet. SKU !='test-001'

 and orddet. SKU !='Sales Tax 1'

 and orddet. SKU !='Coupon'

 and orddet. SKU !='Sales Tax 2'

 and orddet. SKU !='Sales Tax 3'

 and orddet. SKU !='Coupon'

 and orddet. SKU !='Sales Tax 2'

 and orddet. SKU !='Sales Tax 3'

 Group BY inv.PK , cust. CustomerID ORDER BY products\_per\_customer DESC , CustomerID ;

Export the recommendation to CSV using SSIS

Get the users list for which recommendation is to be obtained

select v .CustomerId , v.TotalProductsPurchase ,v .TotalNoOfUniqueProducts FROM

( select CustomerId ,sum (ProductsPerCustomer ) as TotalProductsPurchase ,

count( ProductId) as TotalNoOfUniqueProducts

from SeomRecomSys GROUP BY CustomerId ) v

WHERE --v.TotalProductsPurchase > 1 and

v.TotalNoOfUniqueProducts > 2 ORDER BY v.TotalNoOfUniqueProducts DESC

Convert the query to CSV using SSIS

hadoop fs -put /home/biadmin/Desktop/share/mahout/Input\ Data/seom\ InputData/ByProductId/Get20000CustomersForRecommendation.txt /user/biadmin/InputData/SeomRecommendationEngine/

hadoop fs -put /home/biadmin/Desktop/share/mahout/Input\ Data/seom\ InputData/ByProductId/FullCustomerPurchaseHistory.csv /user/biadmin/InputData/SeomRecommendationEngine/

hadoop fs -rmr /user/biadmin/temp

bin/mahout recommenditembased --input /user/biadmin/InputData/SeomRecommendationEngine/FullCustomerPurchaseHistory.csv --output /user/biadmin/OutputData/FULLSetSEOMRecommendationEngine1121jan01 --usersFile /user/biadmin/InputData/SeomRecommendationEngine/Get20000CustomersForRecommendation.txt -s SIMILARITY\_COSINE --tempDir /user/biadmin/temp/SeomRecommendationEngine0930jan01 --numRecommendations 10

hadoop fs -get /user/biadmin/OutputData/FULLSetSEOMRecommendationEngine1121jan01/part-r-00000 /home/biadmin/Desktop/share/mahout/

Use pig to transform the recommendations to the understandable.

data = load '/home/biadmin/Desktop/share/mahout/RecomResults' using PigStorage(',') as (CustomerId:int,R1:int,P1:chararray,R2:int,P2:chararray,R3:int,P3:chararray,R4:int,P4:chararray,R5:int,P5:chararray,R6:int,P6:chararray,R7:int,P7:chararray,R8:int,P8:chararray,R9:int,P9:chararray,R10:int,P10:chararray);

customerId = foreach data generate CustomerId,R1,P1;

STORE customerId INTO '/home/biadmin/Desktop/share/mahout/PigOutput.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','YES\_MULTILINE','WINDOWS');

customerId = foreach data generate CustomerId,R2,P2;

STORE customerId INTO '/home/biadmin/Desktop/share/mahout/PigOutput2' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','YES\_MULTILINE','WINDOWS');

customerId = foreach data generate CustomerId,R3,P3;

STORE customerId INTO '/home/biadmin/Desktop/share/mahout/PigOutput3' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','YES\_MULTILINE','WINDOWS');

customerId = foreach data generate CustomerId,R4,P4;

STORE customerId INTO '/home/biadmin/Desktop/share/mahout/PigOutput4.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','YES\_MULTILINE','WINDOWS');

customerId = foreach data generate CustomerId,R5,P5;

STORE customerId INTO '/home/biadmin/Desktop/share/mahout/PigOutput5.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','YES\_MULTILINE','WINDOWS');

customerId = foreach data generate CustomerId,R6,P6;

STORE customerId INTO '/home/biadmin/Desktop/share/mahout/PigOutput6.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','YES\_MULTILINE','WINDOWS');

customerId = foreach data generate CustomerId,R7,P7;

STORE customerId INTO '/home/biadmin/Desktop/share/mahout/PigOutput7.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','YES\_MULTILINE','WINDOWS');

customerId = foreach data generate CustomerId,R8,P8;

STORE customerId INTO '/home/biadmin/Desktop/share/mahout/PigOutput8.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','YES\_MULTILINE','WINDOWS');

customerId = foreach data generate CustomerId,R9,P9;

STORE customerId INTO '/home/biadmin/Desktop/share/mahout/PigOutput9.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','YES\_MULTILINE','WINDOWS');

customerId = foreach data generate CustomerId,R10,P10;

STORE customerId INTO '/home/biadmin/Desktop/share/mahout/PigOutput10.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','YES\_MULTILINE','WINDOWS');

CreateTableTOP25000CustomerRecommendations

(CustomerIdint,

ProductNamevarchar(2000),

PurchaseScorefloat

)

select\*INTO#tempfromSEOMRecommendations

recominnerjoininventoryinvoninv.PK=recom.ProductIdORDERBYrecom.CustomerId,recom.PurchaseScoreDESC

INSERTTOP25000CustomerRecommendations

(CustomerId,ProductName,PurchaseScore)

selectCustomerId,ItemName,PurchaseScorefrom#tempORDERBYCustomerId,PurchaseScoreDESC