Table of Contents

[Abstract 1](#_Toc534276173)

[Possible outcomes of the project 1](#_Toc534276174)

[Columns in the data sheets 1](#_Toc534276175)

[Data Description in product\_data.csv 1](#_Toc534276176)

[Data Description in transaction\_data.csv 1](#_Toc534276177)

[List of stores 1](#_Toc534276178)

[User/Customer profile 1](#_Toc534276179)

[Mode of payment 1](#_Toc534276180)

[Gender and distribution 1](#_Toc534276181)

[Age and distribution 1](#_Toc534276182)

[Customer and sales 1](#_Toc534276183)

[Content profile 1](#_Toc534276184)

[Products and units sold 1](#_Toc534276185)

[Products and categories 1](#_Toc534276186)

[Products and prices 1](#_Toc534276187)

[Time series 1](#_Toc534276188)

[Month-wise sales 1](#_Toc534276189)

[Week-wise sales 1](#_Toc534276190)

[Day-wise sales 1](#_Toc534276191)

[Store Profile 1](#_Toc534276192)

[Recommender program with apyori package 1](#_Toc534276193)

[Recommender program with FP Growth package in Scala 1](#_Toc534276194)

[Association Rules OOM with Scala 1](#_Toc534276195)

# Abstract

WIP in a different document.

# Possible outcomes of the project

1. Framework that can be used for any super market
2. Frame work provides insights, predictions and actions required for the business growth for each location
3. Frame work defines automated and manual activities involved, Hardware and software needed
4. Speed and Scalability needs to be taken care
5. Basis : Currently available data and expected growth
6. Separately for brick and mortar and online
7. Consider time series, customer behaviour, available items and location of the store
8. Customer profile : Age, Gender, mode of payment

# Columns in the data sheets

Common columns are in blue

|  |  |
| --- | --- |
| **product\_data.csv** | **transaction\_data.csv** |
| customerID | customerID |
| DOB | DOB |
| Gender | Gender |
| State | State |
| PinCode | PinCode – Customer’s? Store’s? |
| transactionDate | transactionDate |
| store\_code | store\_code |
| store\_description | store\_description |
| till\_no | till\_no |
| transaction\_number\_by\_till | transaction\_number\_by\_till |
| promo\_code | tender\_type |
| promotion\_description | payment\_amount\_by\_tender |
| product\_code | PaymentUsed |
| product\_description |  |
| sale\_price\_after\_promo |  |
| discountUsed |  |

Open items:

1. Identify column(s) that identify unique orders/transactions

# Data Description in product\_data.csv

|  |  |
| --- | --- |
| **Description** | **Value** |
| No. of records in the dataset | 7,981,262 |
| No. of Product units sold | 7,981,262 |
| No. of unique products | 204,115 |
| No. of unique Descriptions | 202,527 |
| No. of unique till numbers | 25 |
| No. of unique customers | 165,055 |
| No. of unique stores | 7 |
| Start date | 01/05/15 |
| End date | 30/06/17 |
| No. of days data available | 791 |

# Data Description in transaction\_data.csv

|  |  |
| --- | --- |
| **Description** | **Value** |
| No. of records in the dataset | 866,277 |
| No. of unique till numbers | 25 |
| No. of unique customers | 164,079 |
| No. of unique stores | 7 |
| Start date | 01/05/15 |
| End date | 30/06/17 |
| No. of days data available | 791 |

Number of all transaction rows: 866277 across all stores

Number of transactions with unique *transaction\_number\_by\_till*: 100052 across all stores

Each*transaction\_number\_by\_till* is occurring multiple times

|  |  |
| --- | --- |
| **transaction\_number\_by\_till** | **Count** |
| 18798 | 32 |
| 19407 | 29 |
| 21093 | 29 |
| 6327 | 28 |
| 43174 | 28 |
| 20316 | 26 |
| 7924 | 26 |
| 24116 | 26 |
| 18714 | 25 |
| 54165 | 25 |

Transaction number *transaction\_number\_by\_till* is not unqiue within the same store and across the stores

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **customerID** | **DOB** | **Gender** | **State** | **PinCode** | **transactionDate** | **store\_code** | **store\_description** | **till\_no** | **tender\_type** | **transaction\_number\_by\_till** | **payment\_amount\_by\_tender** | **PaymentUsed** |
| MMID\_2043103 | 22/05/88 | male | MADHYA PRADESH | 452001 | 04/10/16 | 2906 | MM-INDORE-TREASURE ISLAND | 13 | CASH | 18798 | 95 | CASH |
| MMID\_2047300 | 28/01/72 | male | MADHYA PRADESH | 482002 | 28/05/17 | 2655 | MM-INDORE-MALHAR MEGA MALL | 2 | CASH | 18798 | 5562.6 | CASH |
| MMID\_2048213 | 08/11/54 | male | PUNJAB | 143001 | 19/03/16 | 2615 | MM-AMRITSAR-TRILIUM MALL | 15 | EDCHDFC | 18798 | 2262.5 | HDFC CREDIT CARD |
| MMID\_20419533 | 25/07/60 | male | TAMIL NADU | 625533 | 01/05/16 | 4986 | MM-MADURAI-KOCHADAI VILLAGE | 20 | CASH | 18798 | 1634 | CASH |
| MMID\_204118753 | 25/12/74 | female |  |  | 24/09/16 | 2906 | MM-INDORE-TREASURE ISLAND | 8 | CASH | 18798 | 264.87 | CASH |
| MMID\_20431880 | 22/04/91 | male | GUJARAT | 460001 | 20/12/15 | 2655 | MM-INDORE-MALHAR MEGA MALL | 10 | **EDCUBI** | 18798 | **601** | **UBI CREDIT CARD** |
| MMID\_20431880 | 22/04/91 | male | GUJARAT | 460001 | 20/12/15 | 2655 | MM-INDORE-MALHAR MEGA MALL | 10 | **GVSUPP** | 18798 | **245** | **SUPPLIER GV** |
| MMID\_20431880 | 22/04/91 | male | GUJARAT | 460001 | 20/12/15 | 2655 | MM-INDORE-MALHAR MEGA MALL | 10 | **PYOC** | 18798 | **245** | **Payback** |
| MMID\_20449198 | 14/07/94 | female | MADHYA PRADESH | 452001 | 19/04/17 | 2906 | MM-INDORE-TREASURE ISLAND | 20 | EDCUBI | 18798 | 180 | UBI CREDIT CARD |
| MMID\_20461252 | 02/01/93 | female | MADHYA PRADESH | 452001 | 05/01/17 | 2655 | MM-INDORE-MALHAR MEGA MALL | 4 | EDCICICI | 18798 | 699.8 | ICICI CREDIT CARD |

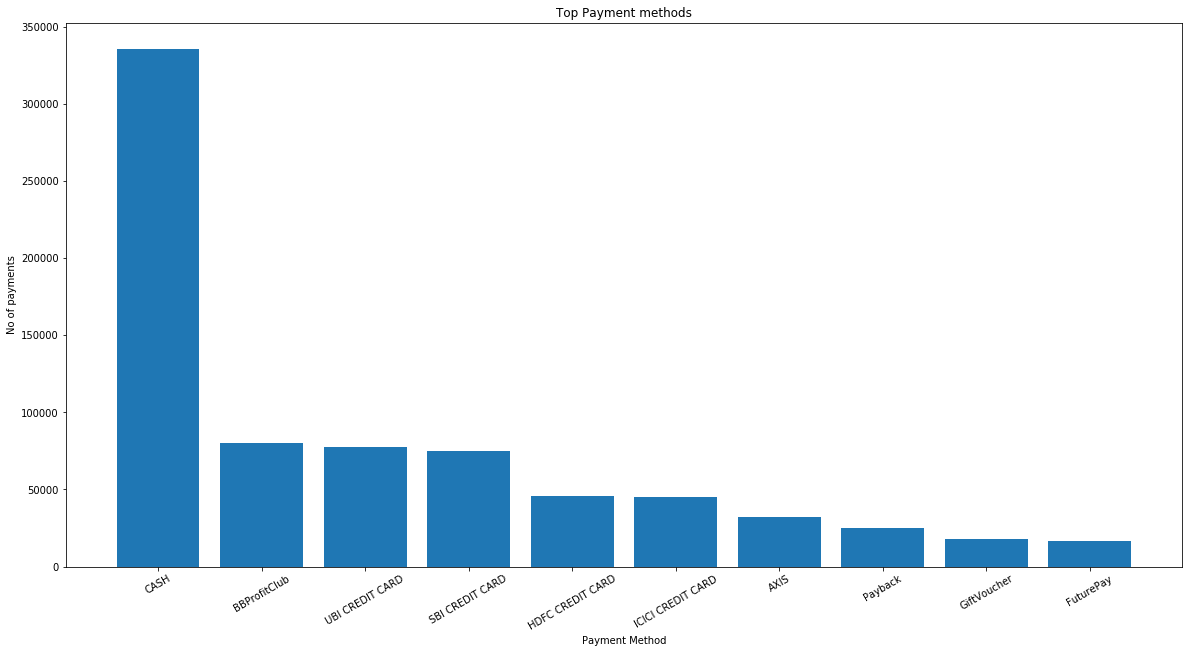
Need to build a new column that identifies unique order.

# List of stores

1. MM-INDORE-MALHAR MEGA MALL
2. MM-LUDHIANA-FEROZEPUR-MF
3. MM-MADURAI-KOCHADAI VILLAGE
4. MM-INDORE-TREASURE ISLAND
5. MM-AMRITSAR-TRILIUM MALL
6. MM-HUBLI-GOKUL ROAD-AKSHAY CEN
7. MM-JAMSHEDPUR-NH-33

# User/Customer profile

## Mode of payment

****

CASH 349375

BBProfitClub 92707

UBI CREDIT CARD 80308

SBI CREDIT CARD 77104

HDFC CREDIT CARD 47144

ICICI CREDIT CARD 46339

AXIS 32426

Payback 25447

GiftVoucher 19738

FuturePay 16942

FGShoppingFest 16861

Cash Back 15960

NaN 7065

FoodCoupon 6995

CREDIT NOTE (ISSUES) 6832

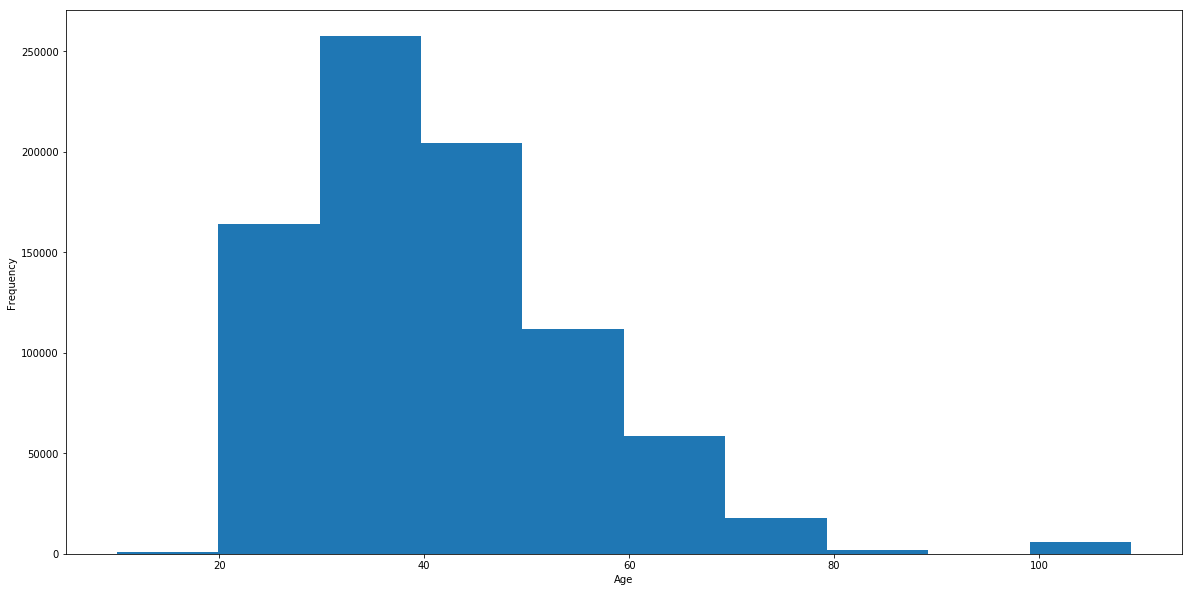
## Gender and distribution



[('male', 582988), ('female', 189630)]

Note: Only transactions that have valid D.O.B value are considered in the above counts

## Age and distribution



## Customer and sales

Customer-wise no. of transactions

|  |  |
| --- | --- |
| **customerID** | **Total payment done** |
| MMID\_211414981 | 32,49,709 |
| MMID\_211419317 | 29,08,563 |
| MMID\_211444893 | 16,66,045 |
| MMID\_20444202 | 13,07,406 |
| MMID\_20453330 | 12,92,226 |

**Customer-wise no. of no. of products (Top 5)**

|  |  |
| --- | --- |
| **customerID** | **Total count of item units bought** |
| MMID\_211421442 | 3284 |
| MMID\_20441636 | 3180 |
| MMID\_20470588 | 2750 |
| MMID\_211419317 | 2594 |
| MMID\_2041210 | 2467 |

|  |  |
| --- | --- |
| **customerID** | **Number of unique txns** |
| MMID\_20444202 | 567 |
| MMID\_20453330 | 389 |
| MMID\_211419317 | 352 |
| MMID\_211414981 | 319 |
| MMID\_2043082 | 257 |

|  |  |
| --- | --- |
| **customerID** | **No. of all txns** |
| MMID\_20444202 | 1065 |
| MMID\_20453330 | 699 |
| MMID\_211419317 | 355 |
| MMID\_20444232 | 335 |
| MMID\_211414981 | 321 |

*Note: When customers buys a basket of items and pays by cash and gift voucher, two entries are made in trasnactions\_data.csv. Hence this table shows more entries than the earlier one.*

# Content profile

## Products and units sold

Top 50 items sold across all the 7 stores

|  |  |  |
| --- | --- | --- |
| **S.No** | **product\_description** | **Units sold** |
| 1 | BB-CB-27X30X208SWG NEW | 60994 |
| 2 | BB-CB-20X25X208SWG NEW | 58437 |
| 3 | SUGAR MEDIUM LOOSE | 51338 |
| 4 | BB-CB-20X25X168SWG-Suitable for ROI New | 41500 |
| 5 | TOMATO LOOSE | 35307 |
| 6 | BB-CB-27X30X168SWG-Suitable for ROI New | 34794 |
| 7 | Fiber bag 45 GSM 20x25 | 31149 |
| 8 | ONION LOOSE | 27963 |
| 9 | TATA SALT PP 1Kg | 27727 |
| 10 | POTATO LOOSE | 25360 |
| 11 | CORIANDER | 24668 |
| 12 | TUR DAL PREM LOOSE | 23638 |
| 13 | BB-CB-16X20X208SWG NEW | 22528 |
| 14 | PB Card FVRL | 21803 |
| 15 | CUCUMBER GREEN LOOSE | 21204 |
| 16 | MAGGI NDL MASALA 420g | 21037 |
| 17 | Carrot English Loose | 20174 |
| 18 | CAPSICUM GREEN | 19628 |
| 19 | LEMON LOOSE | 17676 |
| 20 | LADYFINGER LOOSE | 16987 |
| 21 | BB-CB-16X20X168SWG-Suitable for ROI New | 16491 |
| 22 | DNU | 15738 |
| 23 | GINGER | 15419 |
| 24 | MOONG DAL PREM LOOSE | 15346 |
| 25 | FB SIS NAMKEENS | 14785 |
| 26 | CHANA DAL PREM LOOSE | 14716 |
| 27 | GH DAILY WHO SPICE JEERA100g | 14176 |
| 28 | BOTTLE GOURD LONG | 13902 |
| 29 | AMUL BUTTER 100G | 12792 |
| 30 | MADHUR CRYSTAL SUGAR PP 5Kg | 12629 |
| 31 | GROUNDNUT LOOSE | 12361 |
| 32 | RED LABEL CTC TEA PP 1kg | 12343 |
| 33 | CAULIFLOWER | 12224 |
| 34 | WATERMELON KIRAN | 12093 |
| 35 | VIM UTENSL BAR 3\*200g | 12007 |
| 36 | F'16 FstiveDlghtCrd | 11907 |
| 37 | CABBAGE | 11411 |
| 38 | MINT LEAVES | 11370 |
| 39 | TOMATO COUNTRY | 11131 |
| 40 | SURF EXCEL DET BAR 4\*200G | 11129 |
| 41 | RIN DET BAR ADVANCE 150g | 10775 |
| 42 | GREEN PEAS | 10644 |
| 43 | GH DAILY WHO SPICE JEERA 200g | 10604 |
| 44 | KISSAN KETCHUP DOY PACK 950g | 10460 |
| 45 | CHILLI LIGHT GREEN | 10352 |
| 46 | RIN DET BAR 4\*250g | 10297 |
| 47 | POMEGRANATE REGULAR LOOSE | 10066 |
| 48 | BANANA ROBUSTA RAW | 10009 |
| 49 | TUR DAL ECONOMY LOOSE | 9794 |
| 50 | RIDGE GOURD | 9441 |

Top 20 products across different stores

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **MM-LUDHIANA-FEROZEPUR-MF** |  |  | **MM-HUBLI-GOKUL ROAD-AKSHAY CEN** |  |
| **product\_description** | **count** |  | **product\_description** | **count** |
| ROOH AFZA SHARBAT 750ml | 1347 |  | SUGAR MEDIUM LOOSE | 15455 |
| FORTUNE KACHI GHANI MUSTARD OIL BT 1L | 971 |  | BB-CB-20X25X168SWG-Suitable for ROI New | 13809 |
| CHANA KABULI PREM LOOSE | 945 |  | TUR DAL PREM LOOSE | 12378 |
| MASHAL KACHI GHANI MUSTARD OIL BT 1L | 939 |  | Cloth Bag 20x25 | 8654 |
| FRESH & PURE KCHI GHNI MUSTRD OIL BT 1L | 543 |  | URAD DAL PREM LOOSE | 7737 |
| KURKURE ASSORTED LARGE 3\*100g OP | 503 |  | RUCHI GOLD PALMOLEIN OIL PP 1L | 7696 |
| TTRT FRUIT BASED JUICE LITCHI 2L | 450 |  | GH DAILY WHO SPICE JEERA100g | 6564 |
| GH DAILY PULSES WHO MOONG PRM 1kg | 445 |  | FUTANA FRIED LOOSE | 6558 |
| VERKA GHEE 1 LT TETRA | 439 |  | GROUNDNUT LOOSE | 6372 |
| SOO FRESH CHKN WTH SKIN BROILER LOOSE | 436 |  | MOONG DAL PREM LOOSE | 6170 |
| FB SIS CANDIES | 426 |  | URAD DAL STD LOOSE | 5611 |
| LAYS WFR ASSORTED 2\*95g+MAXX 58g OP | 390 |  | MOONG WHOLE PREM LOOSE | 5500 |
| SOO FRESH COOKD CHKN LEG SPCYTNDR LOOS | 356 |  | BANANA ELAICHI | 5446 |
| TTRT FRUIT BASED JUICE GUAVA 2L | 351 |  | GH DAILY RAVA/SUJI BENSI 1Kg | 5131 |
| SOO FRESH COOKD CHKN TIKKA MALAI LOOSE | 341 |  | CHANA DAL PREM LOOSE | 4877 |
| JIVO CANOLA OIL JR 5L | 313 |  | MADHUR CRYSTAL SUGAR PP 5Kg | 4740 |
| GH RICE BASMATI HYDERABADI BIRYANI 10kg | 271 |  | DHARA SUNFLOWER OIL PP 1L | 4577 |
| KURKURE NAMK SHAHI MIX 1kg | 270 |  | AMUL MALAI PANEER PP 200G | 4271 |
| GH CHANA MAUSAMI 1kg | 269 |  | TUR DAL ECONOMY LOOSE | 4251 |
| NATUREFRESH KCHI GHNI MUSTRD OIL BT 1L | 252 |  | WHEEL DET POW LEMON JASMINE 1Kg | 4103 |
| *[Mustard oil; Kurkure, Lays, Juice, Chicken - Daru babus?]* |  |  | *[Too much sugar]* |  |
|  |  |  |  |  |
| **MM-INDORE-MALHAR MEGA MALL** |  |  | **MM-JAMSHEDPUR-NH-33** |  |
| **product\_description** | **count** |  | **product\_description** | **count** |
| TOMATO LOOSE | 21086 |  | BB-CB-27X30X208SWG NEW | 28015 |
| POTATO LOOSE | 16040 |  | BB-CB-20X25X208SWG NEW | 19112 |
| ONION LOOSE | 15534 |  | BB-CB-16X20X208SWG NEW | 9854 |
| BB-CB-27X30X168SWG-Suitable for ROI New | 14619 |  | MOONG DAL STD LOOSE | 4019 |
| CUCUMBER GREEN LOOSE | 13146 |  | ARWA SONACHUR PREMIUM LOOSE | 3587 |
| CORIANDER | 12319 |  | DHARA KACHI GHANI MUSTARD OIL PP 1L | 3492 |
| FB SIS NAMKEENS | 10827 |  | MASOOR DAL STD LOOSE | 3415 |
| LADYFINGER LOOSE | 9980 |  | VIM UTENSL BAR 125g /130g | 3272 |
| BOTTLE GOURD LONG | 9761 |  | TATA CTC TEA GOLD PP 250g | 2905 |
| Carrot English Loose | 9578 |  | FORTUNE SOYABEAN OIL PP 1L | 2770 |
| CAPSICUM GREEN | 8955 |  | MAAZA NON AERATED 1.2L | 2714 |
| LEMON LOOSE | 8930 |  | SUNFEAST BIS DARK FANTASY CHOCO FILS 75g | 2621 |
| TATA SALT PP 1Kg | 8249 |  | CHANA KABULI DOLLAR LOOSE | 2614 |
| GINGER | 7790 |  | VIM UTENSL BAR 300g | 2549 |
| SOO FRESH BREAD WHITE 400g | 7246 |  | AASHIRVAAD WHOLE WHEAT ATTA PP 5Kg | 2549 |
| RIDGE GOURD | 6694 |  | PRABHUJI SOAN PAPADI ELAICHI BX 500g | 2480 |
| BANANA ROBUSTA RAW | 6614 |  | BRITANNIA BISC MARIE GOLD 2\*300g OP | 2471 |
| BRINJAL BHARTA PURPLE | 6462 |  | ENGINE KACHI GHANI MUSTARD OIL PP 1L | 2391 |
| FB VG CHILLI RED MIRCHI | 6318 |  | NAVRATNA DAL LOOSE | 2387 |
| CABBAGE | 6005 |  | 0TH PULSE CHANA RED LOOSE | 2381 |
| *[Likely that no other good vegetable shops nearby]* |  |  | *[What are the actual items in top 3?]* |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **MM-AMRITSAR-TRILIUM MALL** |  |  | **MM-INDORE-TREASURE ISLAND** |  |
| **product\_description** | **count** |  | **product\_description** | **count** |
| Fiber bag 45 GSM 20x25 | 14781 |  | PB Card FVRL | 5930 |
| MANGO SAFEDA | 1377 |  | SOFRSH VD BREAD WHITE 400g | 3301 |
| RAJMA CHTRA WHITE LOOSE | 1374 |  | SOFRSH VD BREAD WHOLE WHEAT 400g | 1740 |
| BONN BREAD PRIME TIME BROWN PP 400g | 1324 |  | BANANA MADRAS | 1616 |
| Carrot Red | 1123 |  | SOFRSH TRF BREAD MILK 400g | 1445 |
| POTATO SUGARFREE PKD | 1007 |  | SOFRSH TRF BREAD WHOLE WHEAT MLTGR 400g | 1411 |
| BONN BREAD SANDWICH SPECIAL PP 400g | 907 |  | AMUL MILK GOLD MILKY MILK PP 500ml | 1380 |
| MILKFOOD DESI GHEE RF 1L | 826 |  | APPLE SHIMLA PKD | 1186 |
| BONN BREAD PRIME TIME BYRGER PP 225g | 820 |  | SOFRSH TRF BREAD WHITE 400g | 1154 |
| TTRT FRUIT BASED MANGO 2L | 811 |  | SOFRSH TRF MUFFIN VANILLA 6P | 1019 |
| URAD DAL BLACK WHOLE LOOSE | 741 |  | SOFRSH TRF PANEER MALAI 200g | 953 |
| YAKULT FLAVOURED MILK PROBIOTIC BT 5\*65m | 691 |  | SOFRSH TRF MUFFIN CHOCOLATE 6P | 864 |
| TATA CTC TEA PREMIUM PP 1K | 680 |  | APPLE GOLDEN INDIAN LOOSE | 727 |
| PEPSI SOFT DRINK PET 1.5L | 603 |  | SOFRSH TRF PIZZA BASE WHITE 2P | 717 |
| LIMCA SOFT DRINK 1.25L | 579 |  | SOFRSH TRF BREAD BROWN 400g | 687 |
| TATA CTC TEA GOLD PP 1K | 539 |  | SOFY SAN PAD ANTI BACTERA XL 54P OP | 614 |
| FB FR APPLE KINNAUR MEDIUM | 536 |  | MAGGI INST PASTA PAZZTA TOMATO TWIST 64g | 546 |
| MOUNTAIN DEW SOFT DRINK 1.5L | 535 |  | FRESH & PURE TEA KADAK CLASSIC BLEND 1kg | 545 |
| BAMBINO MACARONI PEP ELBOWS 450g | 494 |  | MAGGI INST PASTA PAZZTA MASALA PENNE 65g | 530 |
| APPLE FUJI LOOSE | 493 |  | BOURNVITA CKS CRUNCHIE CHOCOLATEY 120g | 500 |
| *[Why so many fiber bags sold?]* |  |  | *[Working and busy people in the surroundings - Bread, pasta, fruits]* |  |
|  |  |  |  |  |
| **MM-MADURAI-KOCHADAI VILLAGE** |  |  |
| **product\_description** | **count** |  |
| TOMATO COUNTRY | 11131 |  |
| FB VG ONION BUDGET PACK | 7380 |  |
| POMEGRANATE REGULAR LOOSE | 4817 |  |
| FB VG POTATO BUDGET PACK | 4668 |  |
| FB FR BANANA ROBUSTA | 4254 |  |
| COCONUT HUSK LARGE | 4177 |  |
| CHILLI LIGHT GREEN | 4033 |  |
| MANGO BANGANAPALLI | 3844 |  |  |  |
| GOLD WINNER SUNFLOWER OIL PP 1L | 3761 |  |  |  |
| APPLE SHIMLA LOOSE | 3608 |  |  |  |
| URAD DAL GUNDU PRM LOOSE | 3542 |  |  |  |
| ONION ECONOMY PKD | 3407 |  |  |  |
| BRINJAL KATERI VARIKATERI | 2830 |  |  |  |
| CEN BAK SIS BREADS 2 | 2823 |  |  |  |
| ONION SAMBAR | 2651 |  |  |  |
| SUNLAND SUNFLOWER OIL PP 1L | 2522 |  |  |  |
| GH RAW RICE IDLI 5K | 2285 |  |  |  |
| ORANGE NAGPUR PKD | 2260 |  |  |  |
| SURIYAN CHIKKI PEANUT SMALL 100G | 2258 |  |  |  |
| KINNOW LOOSE | 2233 |  |  |  |
| *[A lot of sales in vegetables like in Indore Malhar]* |  |  |  |  |

## Products and categories

Open item: How do we categorize different products across say Home appliances, Vegetables/fruits, Groceries, Fast food items, Cleaning items, décor etc?

## Products and prices

Products with top prices are below. There are errors in entering the data as price of some basic items is marked too high.

|  |  |  |
| --- | --- | --- |
| **S.No** | **product\_description** | **sale\_price\_after\_promo** |
| 1 | Cosmos3 600ml+1000ml+1500ml | 252225 |
| 2 | KORYO 32 Super Slim HDR KLE32EXHN80 | 199485 |
| 3 | Tawa+ Fry Pan Combi Pack | 151696 |
| 4 | Alu Pr Cooker OL - 5Ltr - IB - Favourite | 149410 |
| 5 | SUNLAND SUNFLOWER OIL JR 5L | 114400 |
| 6 | 21Pcs New La Opala D/Set-Blissful Greens | 101250 |
| 7 | 21Pcs New La Opala D/Set-Silken Charm | 101250 |
| 8 | D-Samsung 55K5570 Full HD Smart LED TV | 99911 |
| 9 | HALDIRAM SOANPAPDI PP 500g | 97920 |
| 10 | RAW RICE SONAM LOOSE | 91650 |
| 11 | SAMSUNG 49K6300 CURVED FHD SMART TV | 83999 |
| 12 | D-Samsung 49K5570 Full HD Smart LED TV | 80000 |
| 13 | RICE JEERA KATI LOOSE | 80000 |
| 14 | SONY KDL 50W800D FULL HD ANDROID LED TV | 77990.22 |
| 15 | D-SAMSUNG RS55K5010S9 RFND INOX 591L SBS | 77889.81 |
| 16 | D-Samsung 49K5570 Full HD Smart LED TV | 76010.63 |
| 17 | D-Sony KD-43X8300D 4K HDR Android TV | 73490.01 |
| 18 | OMNI TAWA 250MM+FRY PAN 200MM FREE Alum | 69900 |
| 19 | 21Pcs New La Opala D/Set-Divine Petals | 67500 |
| 20 | 21Pcs New La Opala D/Set-Trinity Green | 67500 |

Some products have negative prices.

Open item : Why are there negative prices?

|  |  |  |
| --- | --- | --- |
| **S.No** | **product\_description** | **sale\_price\_after\_promo** |
| 1 | HUL PUREIT INTELLA STORAGE WATER PURIFIE | -62937 |
| 2 | LG LSA5AF3D 1.5T 3S SAC | -36590 |
| 3 | KORYO 1.5T PORTABLE AC KPA18AF | -28990 |
| 4 | KORYO 1.5T PORTABLE AC KPA18AF | -26991 |
| 5 | Koryo 1.5T 2S Window AC KWR18AO2S | -22290 |
| 6 | AGROFRESH WHOLE WHEAT ATTA PP 10Kg | -19670 |
| 7 | MASOOR DAL STD LOOSE | -13500 |
| 8 | KORYO 32 Super Slim HDR KLE32EXHN80 | -12990 |
| 9 | Koryo 190L DC Refrigerator KDR210DS3F | -11500 |
| 10 | Symphony diet 50i tower cooler | -10491 |
| 11 | GST ROYALE PLUS SCHOTT GLASSTOP GT03L | -7900.75 |
| 12 | BAJAJ TC 2007 TOWER | -6949 |
| 13 | SAMSUNG GW732KD-B/XTL GRILL MWO | -6831.93 |
| 14 | AT Paralite Crimson Red, 79 cm | -6800 |
| 15 | Koryo Healthy Fryer Khf 2215 | -5990 |
| 16 | AT Sky Tracer Dark Slate, 77 cm | -5730 |
| 17 | EKTAA TOOR DAL 1k | -5664 |
| 18 | Preethi 3B Glasstop -Blaze SS | -5599 |
| 19 | Skybags Amore SP Purple 76 cm | -5265 |
| 20 | Skybags Amore SP Purple 76 cm | -5265 |

# Time series

## Month-wise sales

TBD

## Week-wise sales

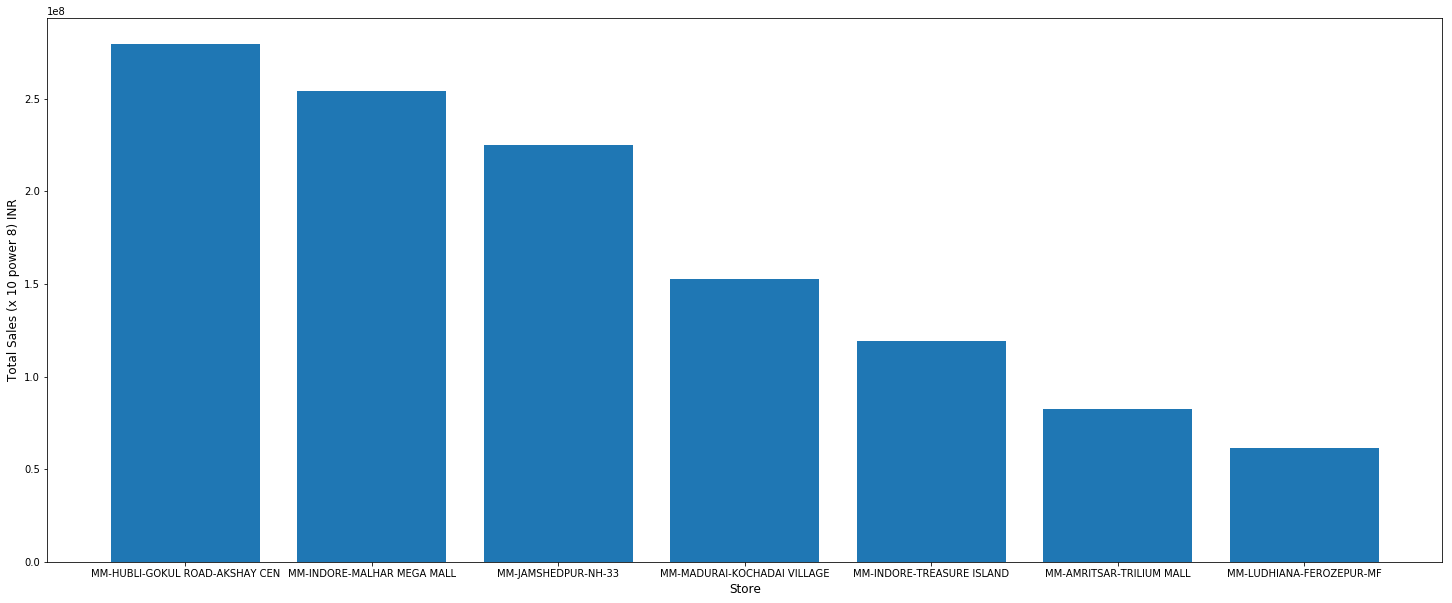
TBD

## Day-wise sales

TBD

# Store Profile

Customer profiles, content profiles and time series data needs to be done at store level as well



|  |  |
| --- | --- |
| **store\_description** | **Total Sales (INR)** |
| MM-HUBLI-GOKUL ROAD-AKSHAY CEN | 27,94,85,800 |
| MM-INDORE-MALHAR MEGA MALL | 25,39,09,200 |
| MM-JAMSHEDPUR-NH-33 | 22,48,64,500 |
| MM-MADURAI-KOCHADAI VILLAGE | 15,26,73,000 |
| MM-INDORE-TREASURE ISLAND | 11,90,41,400 |
| MM-AMRITSAR-TRILIUM MALL | 8,24,50,630 |
| MM-LUDHIANA-FEROZEPUR-MF | 6,13,61,350 |

Top 30 products bought across different stores

| **S.No** | **store** | **product\_description** | **Count** |
| --- | --- | --- | --- |
| **0** | MM-JAMSHEDPUR-NH-33 | BB-CB-27X30X208SWG NEW | 28015 |
| **1** | MM-INDORE-MALHAR MEGA MALL | TOMATO LOOSE | 21086 |
| **2** | MM-JAMSHEDPUR-NH-33 | BB-CB-20X25X208SWG NEW | 19112 |
| **3** | MM-INDORE-MALHAR MEGA MALL | POTATO LOOSE | 16040 |
| **4** | MM-INDORE-MALHAR MEGA MALL | ONION LOOSE | 15534 |
| **5** | MM-HUBLI-GOKUL ROAD-AKSHAY CEN | SUGAR MEDIUM LOOSE | 15455 |
| **8** | MM-AMRITSAR-TRILIUM MALL | Fiber bag 45 GSM 20x25 | 14781 |
| **9** | MM-INDORE-MALHAR MEGA MALL | BB-CB-27X30X168SWG-Suitable for ROI New | 14619 |
| **10** | MM-HUBLI-GOKUL ROAD-AKSHAY CEN | BB-CB-20X25X168SWG-Suitable for ROI New | 13809 |
| **12** | MM-INDORE-MALHAR MEGA MALL | CUCUMBER GREEN LOOSE | 13146 |
| **13** | MM-HUBLI-GOKUL ROAD-AKSHAY CEN | TUR DAL PREM LOOSE | 12378 |
| **14** | MM-INDORE-MALHAR MEGA MALL | CORIANDER | 12319 |
| **18** | MM-MADURAI-KOCHADAI VILLAGE | TOMATO COUNTRY | 11131 |
| **21** | MM-INDORE-MALHAR MEGA MALL | FB SIS NAMKEENS | 10827 |
| **25** | MM-INDORE-MALHAR MEGA MALL | LADYFINGER LOOSE | 9980 |
| **26** | MM-JAMSHEDPUR-NH-33 | BB-CB-16X20X208SWG NEW | 9854 |
| **27** | MM-INDORE-MALHAR MEGA MALL | BOTTLE GOURD LONG | 9761 |
| **28** | MM-INDORE-MALHAR MEGA MALL | Carrot English Loose | 9578 |
| **29** | MM-INDORE-MALHAR MEGA MALL | CAPSICUM GREEN | 8955 |
| **30** | MM-INDORE-MALHAR MEGA MALL | LEMON LOOSE | 8930 |

# Recommender program with apyori package

Apyori/apriori program

*from apyori import apriori*

*association\_rules\_Indore\_Malhar = apriori(records\_Indore\_Malhar, min\_support=0.0045, min\_confidence=0.2, min\_lift=3, min\_length=2)*

*association\_results\_Indore\_Malhar = list(association\_rules\_Indore\_Malhar)*

*for item in association\_results\_Indore\_Malhar:*

*pair = item[0]*

*items = [x for x in pair]*

*list\_of\_rules\_Indore\_Malhar.append([items[0] + " -> " + items[1], str(round(item[1],4)), str(round(item[2][0][2],4)), str(round(item[2][0][3],4))])*

Apyori/apriori Output

|  |  |  |  |
| --- | --- | --- | --- |
| **Rule** | **Confidence** | **Support** | **Lift** |
| ONION LOOSE -> POTATO LOOSE | 0.1178 | 0.6144 | 3.1167 |
| ONION LOOSE -> POTATO LOOSE | 0.0885 | 0.6928 | 3.5148 |
| CORIANDER -> POTATO LOOSE | 0.0672 | 0.8116 | 3.1913 |
| CUCUMBER GREEN LOOSE -> POTATO LOOSE | 0.0666 | 0.8126 | 3.1951 |
| BOTTLE GOURD LONG -> CUCUMBER GREEN LOOSE | 0.0658 | 0.5314 | 3.2287 |
| CUCUMBER GREEN LOOSE -> LADYFINGER LOOSE | 0.0644 | 0.3911 | 3.0777 |
| CUCUMBER GREEN LOOSE -> Carrot English Loose | 0.0644 | 0.391 | 3.1951 |
| ONION LOOSE -> CORIANDER | 0.0638 | 0.8194 | 3.2218 |
| CUCUMBER GREEN LOOSE -> ONION LOOSE | 0.0634 | 0.8064 | 3.1708 |
| CORIANDER -> LADYFINGER LOOSE | 0.0633 | 0.4078 | 3.209 |

# Recommender program with FP Growth package in Scala

Program

val fpg = new FPGrowth()

.setMinSupport(0.0045)

.setNumPartitions(10)

val model = fpg.run(transactions)

model.freqItemsets.collect().foreach { itemset =>

println(itemset.items.mkString("[", ",", "]") + ", " + itemset.freq)

}

# Using this method, we can not get support and lift. Only confidence is available

val minConfidence = 0.5

model.generateAssociationRules(minConfidence).collect().foreach { rule =>

println(

rule.antecedent.mkString("[", ";", "]")

+ " => " + rule.consequent .mkString("[", ",", "]")

+ ", " + rule.confidence)

}

Output

|  |  |
| --- | --- |
| **Rule** | **Confidence** |
| [Nov'16 S.CardBlankt] => [Nov'16 CB Vouchers] | 1 |
| [Aug'16 Scratch Card] => [Aug'16 CB Vouchers] | 1 |
| [Jul'16 Scratch Card] => [July'16 CB Vouchers] | 1 |
| [Oct'16 ScratchCard] => [Oct'16 CB Vouchers] | 0.99844237 |
| [Dec'16 S.Crd-LED TV] => [Dec'16 CB Vouchers] | 0.99268293 |
| [Sep'16 Scratch Card] => [Sep'16 CB Vouchers] | 0.98684211 |
| [Sep'16 CB Vouchers] => [Sep'16 Scratch Card] | 0.98684211 |
| [BRINJAL BHARTA PURPLE;LEMON LOOSE;Carrot English Loose;BOTTLE GOURD LONG;CORIANDER;ONION LOOSE;POTATO LOOSE] => [TOMATO LOOSE] | 0.98028169 |
| [METHI;LEMON LOOSE;Carrot English Loose;BOTTLE GOURD LONG;CORIANDER;ONION LOOSE;POTATO LOOSE] => [TOMATO LOOSE] | 0.97988506 |
| [Aug'16 CB Vouchers] => [Aug'16 Scratch Card] | 0.97706422 |
| [GINGER;CAPSICUM GREEN;LEMON LOOSE;Carrot English Loose;BOTTLE GOURD LONG;CORIANDER;ONION LOOSE] => [TOMATO LOOSE] | 0.97701149 |
| [BRINJAL BHARTA PURPLE;Carrot English Loose;BOTTLE GOURD LONG;LADYFINGER LOOSE;CORIANDER;ONION LOOSE;POTATO LOOSE] => [TOMATO LOOSE] | 0.97668394 |
| [GINGER;LEMON LOOSE;Carrot English Loose;LADYFINGER LOOSE;CORIANDER;CUCUMBER GREEN LOOSE;ONION LOOSE] => [TOMATO LOOSE] | 0.97637795 |
| [METHI;CAPSICUM GREEN;LEMON LOOSE;BOTTLE GOURD LONG;CORIANDER;POTATO LOOSE] => [TOMATO LOOSE] | 0.9744898 |
| [CABBAGE;GINGER;LEMON LOOSE;Carrot English Loose;ONION LOOSE;POTATO LOOSE] => [TOMATO LOOSE] | 0.97435897 |

# Association Rules OOM with [Scala](https://drive.google.com/file/d/1uzRAHJ4o6sBLmBeTn-sJq69RPI1qOApi/view?usp=sharing)

OS: MacOS

With default maxheap of 1GB, the following step runs into OOM. Screenshot is presented after the call stack:

scala> model.freqItemsets.collect().foreach { itemset =>

| println(itemset.items.mkString("[", ",", "]") + ", " + itemset.freq)

| }

*2018-12-30 14:21:37 ERROR SparkUncaughtExceptionHandler:91 - Uncaught exception in thread Thread[Executor task launch worker for task 19,5,main]*

*java.lang.OutOfMemoryError: GC overhead limit exceeded*

*at scala.collection.immutable.List.map(List.scala:288)*

*at org.apache.spark.mllib.fpm.FPGrowth$$anonfun$genFreqItemsets$5.apply(FPGrowth.scala:283)*

*at org.apache.spark.mllib.fpm.FPGrowth$$anonfun$genFreqItemsets$5.apply(FPGrowth.scala:282)*

*at scala.collection.Iterator$$anon$11.next(Iterator.scala:410)*

*at scala.collection.Iterator$class.foreach(Iterator.scala:891)*

*at scala.collection.AbstractIterator.foreach(Iterator.scala:1334)*

*at scala.collection.generic.Growable$class.$plus$plus$eq(Growable.scala:59)*

*at scala.collection.mutable.ArrayBuffer.$plus$plus$eqException in thread "block-manager-slave-async-thread-pool-15" (ArrayBuffer.scala:104)*

*at scala.collection.mutable.ArrayBuffer.$plus$plus$eq(ArrayBuffer.scala:48)*

*at scala.collection.TraversableOnce$class.to(TraversableOnce.scala:310)*

*at scala.collection.AbstractIterator.to(Iterator.scala:1334)*

*at scala.collection.TraversableOnce$class.toBuffer(TraversableOnce.scala:302)*

*at scala.collection.AbstractIterator.toBuffer(Iterator.scala:1334)*

*at scala.collection.TraversableOnce$class.toArray(TraversableOnce.scala:289)*

*at scala.collection.AbstractIterator.toArray(Iterator.scala:1334)*

*at org.apache.spark.rdd.RDD$$anonfun$collect$1$$anonfun$13.apply(RDD.scala:945)*

*at org.apache.spark.rdd.RDD$$anonfun$collect$1$$anonfun$13.apply(RDD.scala:945)*

*at org.apache.spark.SparkContext$$anonfun$runJob$5.apply(SparkContext.scala:2101)*

*at org.apache.spark.SparkContext$$anonfun$runJob$5.apply(SparkContext.scala:2101)*

*at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)*

*at org.apache.spark.java.lang.Error: java.lang.InterruptedException*

*at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1155)*

*at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)*

*at java.lang.Thread.run(Thread.java:748)*

*Caused by: java.lang.InterruptedException*

*at java.lang.Object.wait(Native Method)*

*at java.lang.Object.wait(Object.java:502)*

*at org.apache.spark.storage.BlockInfoManager.lockForWriting(BlockInfoManager.scala:236)*

*at org.apache.spark.storage.BlockManager.removeBlock(BlockManager.scala:1571)*

*at org.apache.spark.storage.BlockManager$$anonfun$removeBroadcast$2.apply(BlockManager.scala:1562)*

*at org.apache.spark.storage.BlockManager$$anonfun$removeBroadcast$2.apply(BlockManager.scala:1562)*

*at scala.collection.Iterator$class.foreach(Iterator.scala:891)*

*at scala.collection.AbstractIterator.foreach(Iterator.scala:1334)*

*at org.apache.spark.storage.BlockManager.removeBroadcast(BlockManager.scala:1562)*

*at org.apache.spark.storage.BlockManagerSlaveEndpoint$$anonfun$receiveAndReply$1$$anonfun$applyOrElse$4.apply$mcI$sp(BlockManagerSlaveEndpoint.scala:66)*

*at org.apache.spark.storage.BlockManagerSlaveEndpoint$$anonfun$receiveAndReply$1$$anonfun$applyOrElse$4.apply(BlockManagerSlaveEndpoint.scala:66)*

*at org.apache.spark.storage.BlockManagerSlaveEndpoint$$anonfun$receiveAndReply$1$$anonfun$applyOrElse$4.apply(BlockManagerSlaveEndpoint.scala:66)*

*scheduler.Task.run(Task.scala:121)*

*at org.apache.spark.executor.Executor$TaskRunner$$anonfun$10.apply(Executor.scala:402)*

*at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1360)*

*at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:408)*

*at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149)*

*at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)*

*at java.lang.Thread.run(Thread.java:748)*

*at org.apache.spark.storage.BlockManagerSlaveEndpoint$$anonfun$1.apply(BlockManagerSlaveEndpoint.scala:86)*

*at scala.concurrent.impl.Future$PromiseCompletingRunnable.liftedTree1$1(Future.scala:24)*

*at scala.concurrent.impl.Future$PromiseCompletingRunnable.run(Future.scala:24)*

*at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149)*

*... 2 more*

*2018-12-30 14:21:37 WARN JavaUtils:113 - Attempt to delete using native Unix OS command failed for path = /private/var/folders/48/fh0qdcz15cs8prg4qb0qr3s40000gn/T/spark-e01899b9-dda1-4c0c-a706-3c351aa36d33/repl-ca39df2d-09f6-42e9-a2db-ae8015781832. Falling back to Java IO way*

*java.io.IOException: Failed to delete: /private/var/folders/48/fh0qdcz15cs8prg4qb0qr3s40000gn/T/spark-e01899b9-dda1-4c0c-a706-3c351aa36d33/repl-ca39df2d-09f6-42e9-a2db-ae8015781832*

*at org.apache.spark.network.util.JavaUtils.deleteRecursivelyUsingUnixNative(JavaUtils.java:171)*

*at org.apache.spark.network.util.JavaUtils.deleteRecursively(JavaUtils.java:110)*

*at org.apache.spark.network.util.JavaUtils.deleteRecursively(JavaUtils.java:91)*

*at org.apache.spark.util.Utils$.deleteRecursively(Utils.scala:1062)*

*at org.apache.spark.util.ShutdownHookManager$$anonfun$1$$anonfun$apply$mcV$sp$3.apply(ShutdownHookManager.scala:65)*

*at org.apache.spark.util.ShutdownHookManager$$anonfun$1$$anonfun$apply$mcV$sp$3.apply(ShutdownHookManager.scala:62)*

*at scala.collection.IndexedSeqOptimized$class.foreach(IndexedSeqOptimized.scala:33)*

*at scala.collection.mutable.ArrayOps$ofRef.foreach(ArrayOps.scala:186)*

*at org.apache.spark.util.ShutdownHookManager$$anonfun$1.apply$mcV$sp(ShutdownHookManager.scala:62)*

*at org.apache.spark.util.SparkShutdownHook.run(ShutdownHookManager.scala:216)*

*at org.apache.spark.util.SparkShutdownHookManager$$anonfun$runAll$1$$anonfun$apply$mcV$sp$1.apply$mcV$sp(ShutdownHookManager.scala:188)*

*at org.apache.spark.util.SparkShutdownHookManager$$anonfun$runAll$1$$anonfun$apply$mcV$sp$1.apply(ShutdownHookManager.scala:188)*

*at org.apache.spark.util.SparkShutdownHookManager$$anonfun$runAll$1$$anonfun$apply$mcV$sp$1.apply(ShutdownHookManager.scala:188)*

*at org.apache.spark.util.Utils$.logUncaughtExceptions(Utils.scala:1945)*

*at org.apache.spark.util.SparkShutdownHookManager$$anonfun$runAll$1.apply$mcV$sp(ShutdownHookManager.scala:188)*

*at org.apache.spark.util.SparkShutdownHookManager$$anonfun$runAll$1.apply(ShutdownHookManager.scala:188)*

*at org.apache.spark.util.SparkShutdownHookManager$$anonfun$runAll$1.apply(ShutdownHookManager.scala:188)*

*at scala.util.Try$.apply(Try.scala:192)*

*at org.apache.spark.util.SparkShutdownHookManager.runAll(ShutdownHookManager.scala:188)*

*at org.apache.spark.util.SparkShutdownHookManager$$anon$2.run(ShutdownHookManager.scala:178)*

*at org.apache.hadoop.util.ShutdownHookManager$1.run(ShutdownHookManager.java:54)*

*2018-12-30 14:21:37 ERROR TaskSchedulerImpl:70 - Ignoring update with state FAILED for TID 19 because its task set is gone (this is likely the result of receiving duplicate task finished status updates) or its executor has been marked as failed.*

*org.apache.spark.SparkException: Job aborted due to stage failure: Task 7 in stage 4.0 failed 1 times, most recent failure: Lost task 7.0 in stage 4.0 (TID 23, localhost, executor driver): ExecutorLostFailure (executor driver exited caused by one of the running tasks) Reason: Executor heartbeat timed out after 162069 ms*

*Driver stacktrace:*

*at org.apache.spark.scheduler.DAGScheduler.org$apache$spark$scheduler$DAGScheduler$$failJobAndIndependentStages(DAGScheduler.scala:1887)*

*at org.apache.spark.scheduler.DAGScheduler$$anonfun$abortStage$1.apply(DAGScheduler.scala:1875)*

*at org.apache.spark.scheduler.DAGScheduler$$anonfun$abortStage$1.apply(DAGScheduler.scala:1874)*

*at scala.collection.mutable.ResizableArray$class.foreach(ResizableArray.scala:59)*

*at scala.collection.mutable.ArrayBuffer.foreach(ArrayBuffer.scala:48)*

*at org.apache.spark.scheduler.DAGScheduler.abortStage(DAGScheduler.scala:1874)*

*at org.apache.spark.scheduler.DAGScheduler$$anonfun$handleTaskSetFailed$1.apply(DAGScheduler.scala:926)*

*at org.apache.spark.scheduler.DAGScheduler$$anonfun$handleTaskSetFailed$1.apply(DAGScheduler.scala:926)*

*at scala.Option.foreach(Option.scala:257)*

*at org.apache.spark.scheduler.DAGScheduler.handleTaskSetFailed(DAGScheduler.scala:926)*

*at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.doOnReceive(DAGScheduler.scala:2108)*

*at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.onReceive(DAGScheduler.scala:2057)*

*at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.onReceive(DAGScheduler.scala:2046)*

*at org.apache.spark.util.EventLoop$$anon$1.run(EventLoop.scala:49)*

*at org.apache.spark.scheduler.DAGScheduler.runJob(DAGScheduler.scala:737)*

*at org.apache.spark.SparkContext.runJob(SparkContext.scala:2061)*

*at org.apache.spark.SparkContext.runJob(SparkContext.scala:2082)*

*at org.apache.spark.SparkContext.runJob(SparkContext.scala:2101)*

*at org.apache.spark.SparkContext.runJob(SparkContext.scala:2126)*

*at org.apache.spark.rdd.RDD$$anonfun$collect$1.apply(RDD.scala:945)*

*at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope.scala:151)*

*at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope.scala:112)*

*at org.apache.spark.rdd.RDD.withScope(RDD.scala:363)*

*at org.apache.spark.rdd.RDD.collect(RDD.scala:944)*

*... 51 elided*