

Data Analysis on Socio-Economic and Well-being Fitness

Objective:

The project is based on Data Analysis for Economy, Seasonality, product selling, well-being Fitness, Financial exchange, Sports interest and Education class and age breakdown. This project would help in analyzing socio-economic status, their wellbeing fitness in for each age group and various period.

Given Data:

Sports & Fitness Equipment sales details and Customer details

| Trans id | Date | Cust id | Amount | Category | Item name | City | State | Mode of payment |
|----------|------|---------|--------|----------|-----------|------|-------|-----------------|
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |

| Cust id | First Name | Last Name | Age | Profession |
|---------|------------|-----------|-----|------------|
| | | | | |
| | | | | |
| | | | | |

Analysis on following factors:

1. Economic Status
2. Seasonality
3. Financial Exchanges
4. Healthcare
5. Sports Awareness

1. Economic Status:

Economic Analysis is powerful way of thinking about business world. There is phenomenal growth of e-commerce is increased access to the internet. Financial Services organizations are using data mined from customer interactions to slice and dice their users into finely tuned segments. It also tells us how much individual, single organization or government using e-commerce.

For Example: Analyzing Economic status in for each age group and each profession on citizenship i.e. finding out the sales happened for person with age groups and profession.

We have achieved the same in our project in various Tasks mentioned below.

Task1: Finding all the transaction where Amount>160.

Task2: Counting all the transaction where amount is between 175 and 200.

Task3: Calculating the total sum, total count, average of all the transaction for each user id.

2. Seasonality

We know that retail sales move with the seasons, Data is very predictable. Seasonality is a characteristic of a time series in which the data experiences regular and predictable changes that recur every calendar year. Any predictable change or pattern in a time series that recurs or repeats over a one-year period can be said to be seasonal. Retail sales in general always rise in the fourth quarter, but e-commerce sales do so even more intensely.

For Example: Finding the period of time with maximum sales.

We have achieved the same in our project in various Tasks mentioned below.

Task 4: Calculating total sales amount for each Month.

Task 5: Dividing the file into 12 files, where each file containing each month of data. For ex: file 1 should contain data of January transactions, file 2 should contain data of February transactions and soon.

Task 9: Finding the user who has spent the max amount in July month.

3. Financial Exchanges:

E-Commerce or Electronics Commerce sites use case and electronic payment where electronic payment refers to paperless monetary transactions. Electronic payment has revolutionized the business processing by various factors. This enables these financial institutions to create increasingly relevant and sophisticated offers. Being user friendly and less time consuming than manual processing, helps business organization to expand its market reach / expansion.

For Example: Finding out the persons who made their purchases on various modes of payment by cash, credit card, debit card, Internet Banking NEFT.

We have achieved the same in our project in various Tasks mentioned below.

Task 6: Sorting the whole file on the basis of amt.

Task 7: Finding the name of top 3 spenders.

Task 8: Finding the name of user who has spent the maximum amount.

Future Enhancement: Finding the number of persons who made their purchase based on mode of payment.

4. Healthcare

We have to focus on Healthcare - both in terms of fitness and diseases outbreak.

Healthcare organizations are turning to data analysis to improve their care delivery and the cost of care. This services to this population to help fight the obesity.

EX: identifying No of persons finding their interest in sports, gyms, yoga exc.

Future Enhancement:

We can implement the same in our project in various Tasks likes finding the total number of persons based on category of purchase. Grouping the data based on Exercise & Fitness and Gymnastics.

5. Sports Awareness:

Sports and games form an essential part of human resource Development. Sport development is a national priority, as it promotes active Lifestyle, child and youth

development, social inclusiveness, Employment opportunities, peace and development, and above all a Sense of belongingness and national pride.

For example: Finding the number of person who is interest in various sports.

Future Enhancement:

We can implement the same in our project in various Tasks likes finding the total number of persons who were interest in various sports, we can group the persons in based on their interest. Also we can find the maximum persons who were interested in particular sports.

Technology used:

Apache Hadoop:

Apache Hadoop an open-source software framework used for distributed storage and processing of very large data sets. It consists of computer clusters built from commodity hardware.

Characteristics of Hadoop:


- Reliable
- Flexible
- Economical
- Scalable

Map reduce java coding

Hadoop MapReduce is a software framework for easily writing applications which process vast amounts of data (multi-terabyte data-sets) in-parallel on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner.

A MapReduce job usually splits the input data-set into independent chunks which are processed by the map tasks in a completely parallel manner. The framework sorts the outputs of the maps, which are then input to the reduce tasks. Typically both the input and the output of the job are stored in a file-system. The framework takes care of scheduling tasks, monitoring them and re-executes the failed tasks.

Generating userid whose transaction amount is greater than 190.

 hduser@ubuntu64server: ~

```
user id--->4005468---Amount---> 197.26
user id--->4004910---Amount---> 192.26
user id--->4005878---Amount---> 195.11
user id--->4006098---Amount---> 199.49
user id--->4008830---Amount---> 196.2
user id--->4008320---Amount---> 190.36
user id--->4006437---Amount---> 196.66
user id--->4005415---Amount---> 195.13
user id--->4005158---Amount---> 194.03
user id--->4002147---Amount---> 196.31
user id--->4000309---Amount---> 192.39
user id--->4003684---Amount---> 190.21
user id--->4005411---Amount---> 196.69
user id--->4005712---Amount---> 194.96
user id--->4005522---Amount---> 194.7
user id--->4000456---Amount---> 196.72
user id--->4006181---Amount---> 192.1
user id--->4003750---Amount---> 199.93
user id--->4002759---Amount---> 195.77
user id--->4006192---Amount---> 191.37
user id--->4000817---Amount---> 195.83
user id--->4000889---Amount---> 195.33
user id--->4004072---Amount---> 198.57
user id--->4009725---Amount---> 196.32
user id--->4007024---Amount---> 198.25
user id--->4004486---Amount---> 191.94
user id--->4003329---Amount---> 193.5
user id--->4007255---Amount---> 199.16
user id--->4006905---Amount---> 193.83
user id--->4008851---Amount---> 197.9
user id--->4001337---Amount---> 195.79
user id--->4001714---Amount---> 193.66
user id--->4003161---Amount---> 194.15
user id--->4003791---Amount---> 196.39
user id--->4007751---Amount---> 196.01
user id--->4008608---Amount---> 198.94
user id--->4004646---Amount---> 197.07
user id--->4006836---Amount---> 191.32
user id--->4007935---Amount---> 190.24
user id--->4005076---Amount---> 199.92
user id--->4000927---Amount---> 196.09
user id--->4000172---Amount---> 198.33
user id--->4000454---Amount---> 199.35
user id--->4000978---Amount---> 199.39
user id--->4004939---Amount---> 198.32
user id--->4008449---Amount---> 192.67
user id--->4004318---Amount---> 199.07
user id--->4008637---Amount---> 198.4
user id--->4003685---Amount---> 191.29
```

Total sum, count and average amount of sale of particular user:

```
hduser@ubuntu64server:~$ hadoop fs -cat /p12/p*  
user id-->4000400--sum--> 1128.7-->count 11--average--> 102.60909090909091  
hduser@ubuntu64server:~$
```

Total sales of user of each month

hduser@ubuntu64server: ~

```
12 20.4  
12 180.07  
12 142.84  
12 53.04  
12 58.34  
12 198.64  
12 118.23  
12 153.58  
12 193.82  
12 191.47  
12 75.24  
12 20.58  
12 46.81  
12 68.44  
12 106.16  
12 124.39  
12 113.0  
12 100.94  
12 167.44  
12 50.43  
12 97.08  
12 24.31  
12 127.69  
12 198.72  
12 32.21  
12 21.41  
12 95.81  
12 21.43  
12 114.95  
12 175.29  
12 47.06  
12 60.17  
12 195.13  
12 94.9  
12 56.76  
12 197.48  
12 152.51  
12 195.4  
12 5.06  
12 22.16  
12 30.34  
12 17.67  
12 150.23  
12 157.43  
12 126.58  
12 110.83  
12 138.43  
12 172.67
```

Record of Feb month:

hduser@ubuntu64server: ~

| | | | | | | |
|------------------|---------------------|---------------|------------|--------|------------------|----------|
| 2 | 00025725 | 02-10-2011 | 4000702 | 144.86 | Outdoor Recreati | |
| on | Motorsports | Santa Ana | California | credit | | |
| 2 | 00006644 | 02-20-2011 | 4005475 | 055.91 | Jumping | T |
| rampolines | Coral Springs | Florida | | credit | | |
| 2 | 00002506 | 02-19-2011 | 4009147 | 192.01 | Outdoor Recreati | |
| on | Track & Field | Coral Springs | Florida | credit | | |
| 2 | 00034623 | 02-16-2011 | 4008634 | 051.85 | Games | Portable |
| | Electronic Games | Omaha | Nebraska | credit | | |
| 2 | 00025732 | 02-09-2011 | 4004961 | 074.93 | Water Sports | L |
| life Jackets | New Orleans | Louisiana | | credit | | |
| 2 | 00048437 | 02-17-2011 | 4006557 | 135.36 | Water Sports | B |
| oating | San Antonio | Texas | credit | | | |
| 2 | 00001237 | 02-18-2011 | 4004161 | 198.49 | Jumping | P |
| ogo Sticks | New York | New York | | credit | | |
| 2 | 00042569 | 02-19-2011 | 4001706 | 073.89 | Outdoor Recreati | |
| on | Motorsports | Alexandria | Virginia | credit | | |
| 2 | 00031766 | 02-07-2011 | 4004555 | 018.80 | Outdoor Recreati | |
| on | Tetherball | New Orleans | Louisiana | cash | | |
| 2 | 00008880 | 02-09-2011 | 4002573 | 061.08 | Water Sports | W |
| ater Tubing | Salem | Oregon | credit | | | |
| 2 | 00006637 | 02-11-2011 | 4009305 | 132.47 | Jumping | B |
| ungee Jumping | Houston | Texas | credit | | | |
| 2 | 00005609 | 02-12-2011 | 4009523 | 117.59 | Racquet Sports | B |
| adminton | Phoenix | Arizona | | credit | | |
| 2 | 00042568 | 02-08-2011 | 4005165 | 023.84 | Water Sports | B |
| oating | Madison | Wisconsin | cash | | | |
| 2 | 00011776 | 02-06-2011 | 4001285 | 054.07 | Games | Dominoes |
| | Lexington | Kentucky | credit | | | |
| 2 | 00034639 | 02-25-2011 | 4003008 | 112.48 | Exercise & Fitne | |
| ss | Yoga & Pilates | Coral Springs | Florida | credit | | |
| 2 | 00002088 | 02-09-2011 | 4005864 | 156.70 | Outdoor Recreati | |
| on | Shooting Games | Scottsdale | Arizona | credit | | |
| 2 | 00012290 | 02-09-2011 | 4006619 | 033.50 | Exercise & Fitne | |
| ss | Abdominal Equipment | El Paso | Texas | cash | | |
| 2 | 00025759 | 02-09-2011 | 4009104 | 192.92 | Team Sports | S |
| occer | Springfield | Illinois | credit | | | |
| 2 | 00044314 | 02-19-2011 | 4000363 | 150.46 | Exercise & Fitne | |
| ss | Medicine Balls | Fremont | California | credit | | |
| 2 | 00025767 | 02-13-2011 | 4004739 | 129.48 | Team Sports | I |
| ndoor Volleyball | Seattle | Washington | credit | | | |
| 2 | 00047962 | 02-01-2011 | 4007952 | 123.56 | Winter Sports | S |
| nowshoeing | Hampton | Virginia | | credit | | |
| 2 | 00034661 | 02-04-2011 | 4007960 | 150.26 | Gymnastics | G |
| ymnastics Rings | Portland | Oregon | credit | | | |
| 2 | 00008868 | 02-20-2011 | 4005073 | 107.30 | Exercise & Fitne | |
| ss | Weight Benches | Houston | Texas | credit | | |

Total Sales amount of each month:

```
hduser@ubuntu64server:~$ hadoop fs -cat /p12/p*
08      434255.0100000014
```

Conclusion:

Data Analysis in E-Commerce can enhance various factors in terms of sales, transaction, communication with the customers. The project analyzed the status of Economy, Seasonality, product selling, well-being Fitness, Financial exchange, Sports interest on factors like each age group and various period.