**Data Analysis on Socio-Economic and Well-being Healthcare**

Objective:

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

**Date: 22/11/2016**

**Prepared by: Sangeetha**

**Given Data: Sports & Fitness Equipment sales**

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

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

**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.

**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

**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.

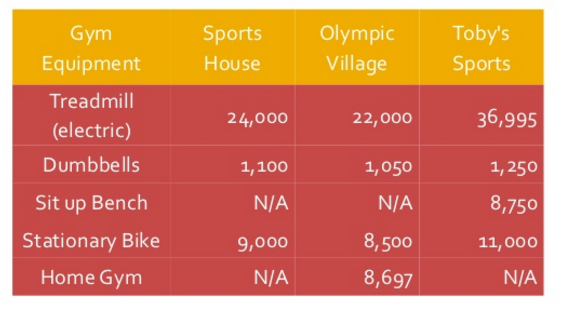
For Example: Finding out the persons who made their purchases on various modes of payment by cash, credit card, debit card, Internet Banking NEFT.

**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.



**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.

**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

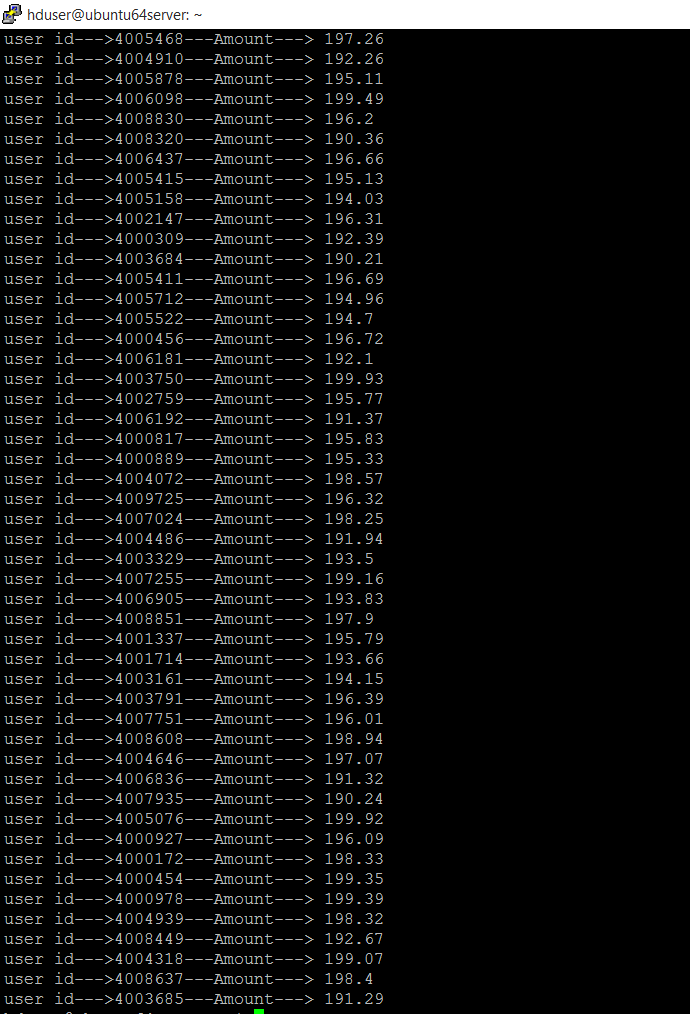
Scalable

**Map reduce java coding**

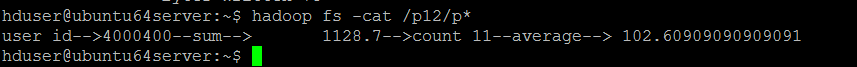
Hadoop MapReduce is a software framework for easily writing applications which process vast amounts of data (multi-terabyte data-sets) in-parallel on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner.

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.

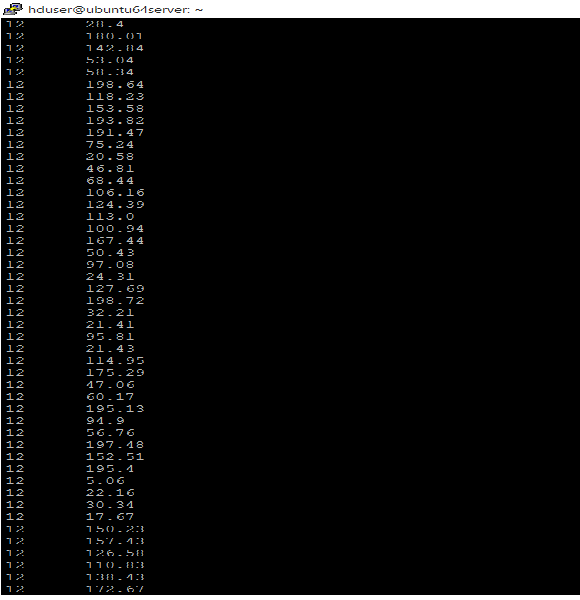
**Generating userid whose transaction amount is greater than 190.**

****

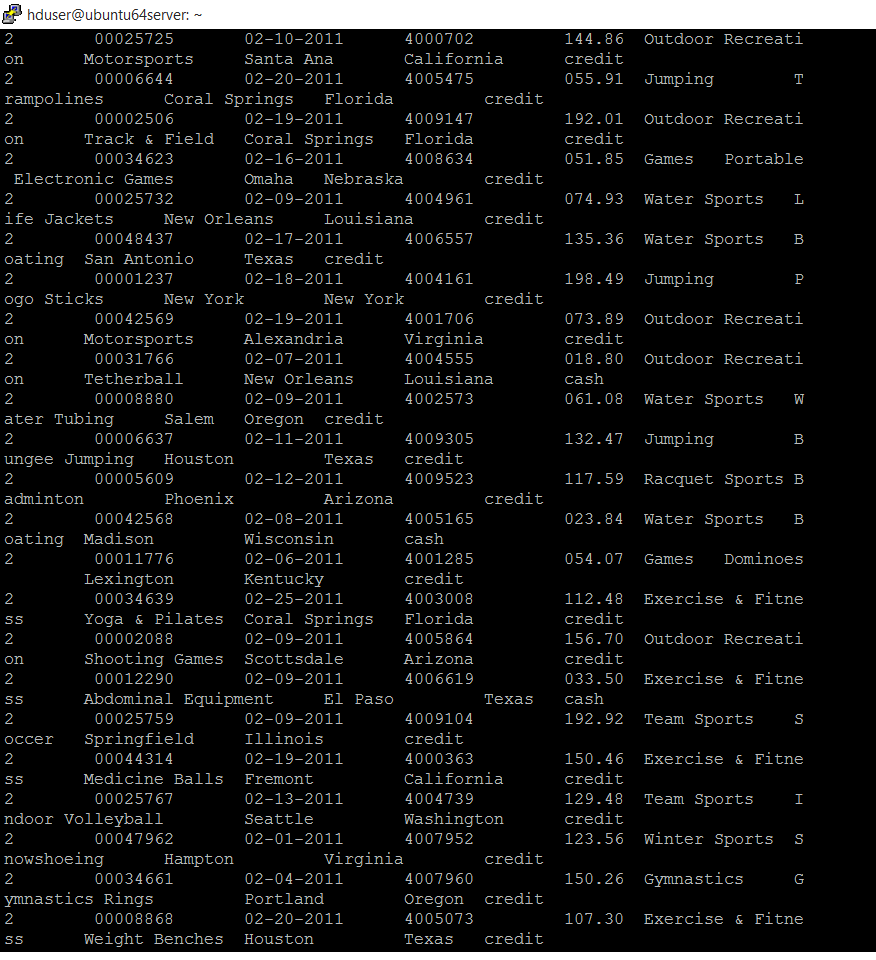
**Total sum, count and average amount of sale of particular user:**

****

**Total sales of user of each month**

****

**Record of Feb month:**

****