SCHOOL OF ENGINEERING AND TECHNOLOGY
DEPARTMENT OF
COMPUTER SCIENCE AND ENGINEERING

CIA 3 - Mini project

REPORT ON - Dataset: Supply chain and logistics Managment

Submitted by:
Batch-9
Praveen kumar C-2062254
Adhil A-2062201
Yukeskaanth V-2062212

Subject In-Charge
Dr. Sathish Kumar

Department of Computer Science and Engineering

School of Engineering and Technology,
CHRIST (Deemed to Be University), Kumbalagodu,
Bangalore - 560 074.
April 2023

INDEX

OBJECTIVES

1. Display the order id with has the
   weight of less than 100
2. List out the carrier or order id which
   goes through Ground Transporations
   with mininum cost
3. To find the average days a customer has waited for the shipment to arrive.

INTRODUCTION

Supply chain management refers to the coordination and management of all the activities involved in the production and delivery of goods and services, from the procurement of raw materials to the delivery of finished products to customers. It encompasses all the processes, people, and systems involved in getting a product from the manufacturer to the customer. This includes managing suppliers, production facilities, warehouses, transportation, inventory management, and customer service.

Logistics management, on the other hand, is a subset of supply chain management that focuses specifically on the planning, implementation, and control of the movement of goods and materials within a supply chain. It involves the management of transportation, warehousing, and inventory to ensure that products are delivered to the right place, at the right time, and in the right condition. Logistics management involves the optimization of the supply chain to reduce costs, improve efficiency, and increase customer satisfaction.

In summary, supply chain management is concerned with the overall management of the supply chain, while logistics management is concerned with the specific management of the movement and storage of goods within the supply chain. Both are critical to the success of any business that produces or delivers products to customers.

The next stage of the analysis will be exploratory data analysis. This stage involves using descriptive statistics and data visualization techniques to explore the data and gain insights into the patterns and trends that emerge.

We will begin by using descriptive statistics to summarize the data and identify any outliers or unusual values. We will also calculate measures of central tendency, such as the mean and median, and measures of dispersion, such as the standard deviation and range, to gain a deeper understanding of the distribution of the data.

We will then use data visualization techniques, such as scatter plots, box plots, and histograms, to explore the relationships between different variables and identify any patterns or trends.

Hypothesis Testing:

The next stage of the analysis will involve hypothesis testing. This stage involves using statistical methods to test hypotheses and make inferences about the data.

Data Visualization:

The final stage of the analysis will be data visualization. This stage involves creating interactive visualizations and dashboards to help us better understand the data and communicate our findings to others. We may create a dashboard to Track the individual product in the supply chain , allowing users to filter the data by order id,minimum cost or other variables. We may also create interactive maps to explore the geographic distribution of supply chain and their performance in different regions of theworld.

DESCRIPTION OF THE DATASET

Dataset is divided into 7 tables, one table for all orders that needs to be assigned a route – OrderList table, and 6 additional files specifying the problem and restrictions. For instance, the FreightRates table describes all available couriers, the weight gaps for each individual lane and rates associated. The PlantPorts table describes the allowed links between the warehouses and shipping ports in real world. Furthermore, the ProductsPerPlant table lists all supported warehouse-product combinations. The VmiCustomers lists all special cases, where warehouse is only allowed to support specific customer, while any other non-listed warehouse can supply any customer. Moreover, the WhCapacities lists warehouse capacities measured in number of orders per day and the WhCosts specifies the cost associated in storing the products in given warehouse measured in dollars per unit.

Contents of the Dataset:

Supply chain management (SCM) involves the coordination and management of all activities involved in the production and delivery of goods or services, from the sourcing of raw materials to the final delivery of the finished product to the end customer. Logistics management, on the other hand, refers to the planning, implementation, and control of the flow of goods, services, and information between the point of origin and the point of consumption.

1.Order ID: A unique identifier for each order placed.
2.Customer ID: A unique identifier for each customer.
3.Product ID: A unique identifier for each product.
4.Quantity: The number of units of the product ordered.
5.Order Date: The date on which the order was placed.
6.Shipment Date: The date on which the product was shipped.
7.Delivery Date: The date on which the product was delivered to the customer.
8.Supplier ID: A unique identifier for each supplier.
9.Supplier Name: The name of the supplier.
10.Raw Material ID: A unique identifier for each raw material used in the production process.
11.Raw Material Name: The name of the raw material.
12.Raw Material Quantity: The quantity of the raw material used in the production process.
13.Production Date: The date on which the product was produced

PREPROCESSING STEPS USED IN DATASET

Preprocessing is a crucial step in data analysis as it helps to clean and transform the raw data into a more usable and meaningful format for analysis. In this report, we will discuss some of the most commonly used preprocessing techniques in data analysis.

The Supply chain and logistics management dataset on Kaggle is a comprehensive collection of data on every product, Carrier Type and Mode of transport performance in the supply chain management . Preprocessing is an essential steping preparing this data for analysis, as it involves cleaning, transforming, and formatting the data to make it moresuitable for analysis. In this report, we will discuss some of the preprocessing techniques used in this dataset.

Data Cleaning:

The first step in preprocessing the Supply chain and Logistics Managament dataset is data cleaning. This involves identifying and correcting errors, inconsistencies, and missing values in the data. Some of the common cleaning techniques usedin this dataset include:

- Removing duplicates: Some entries in the dataset may be duplicates of other entries, which can skew the results of the analysis. These duplicates can be identified and removed using various techniques, such as sorting the data by key attributes and comparing adjacent entries.
- Handling missing values: The dataset may contain missing values for certain attributes, such as player positions or team formations. These missing values can be filled in using various techniques, such as imputation, interpolation, or deletion.
- Correcting errors: The dataset may contain errors or inconsistencies in the data, such as misspelled player names or incorrect match scores. These errors can be corrected using various techniques, such as manual correction or automated algorithms.

Data Transformation:

Once the data has been cleaned, the next step is data transformation. This involves converting the data into a format that is more suitable for analysis. Some of the common transformation techniques used in this dataset include:

- Normalization: The data may be normalized to ensure that it is on a consistent scale. This can be achieved using various techniques, such as z-score normalization or min-max scaling.
- Aggregation: The data may be aggregated to summarize the performance of teams or players over a certain period. This can be achieved using various techniques, such as averaging, summation, or maximum/minimum values.
- Feature selection: The dataset may contain a large number of features that are not relevant to the analysis. Feature selection involves identifying and selecting the most relevant features for the analysis.

Data Formatting

The final step in preprocessing the Supply chain and Logistics Managament dataset is data formatting. This involves formatting the datainto a structure that is suitable for analysis. Some of the common formatting techniques used in this dataset include:

- Reshaping the data: The data may be reshaped to convert it from a wide format to a long format or vice versa. This can be achieved using various techniques, such as the pivot function or the melt function.
- Encoding categorical data: The data may contain categorical data, such as team names or player positions,that need to be encoded into numerical values. This can be achieved using various techniques, such as one-hot encoding or label encoding.
- Splitting the data: The dataset may be split into training and testing sets to evaluate the performance of predictive models. This can be achieved using various techniques, such as random sampling or stratified sampling.

Overall, the preprocessing techniques used in the Supply chain and Logistics Managament dataset on Kaggle are essential for preparing the data for analysis. By cleaning, transforming, and formatting the data, researchers and analysts can ensure that the data is accurate, relevant, and suitable for the analysis.

With these preprocessing techniques in place, the data can be analyzed to identify patterns and trends in team andplayer performance, as well as the evolution of the sport over time.

CODING

Main function:

```
import java.util.Scanner;

public class Driver {

  public static void main(String[] args) throws Exception {
    Scanner scanner = new Scanner(System.in);

    System.out.println("Please enter 1 or 2:");
    int input = scanner.nextInt();

    switch (input) {
      case 1:
        OrderWeightLessThan100.run();
        break;
      case 2:
        DelayedOrders.run();
        break;
      default:
        System.out.println("Invalid input. Please enter 1 or 2.");
        break;
    }

    scanner.close();
  }

  public static void function1() {
```

```
  // implementation of function1
  System.out.println("Executing function1");
 }

 public static void function2() {
  // implementation of function2
  System.out.println("Executing function2");
 }
}
```

Objective 1

Mapper and reducer Function:

```
 import java.io.IOException;
 import java.util.StringTokenizer;
 import org.apache.hadoop.conf.Configuration;
 import org.apache.hadoop.fs.Path;
 import org.apache.hadoop.io.DoubleWritable;
 import org.apache.hadoop.io.Text;
 import org.apache.hadoop.mapreduce.Job;
 import org.apache.hadoop.mapreduce.Mapper;
 import org.apache.hadoop.mapreduce.Reducer;
 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

 public class OrderWeightLessThan100 {

  public static class OrderMapper
      extends Mapper<Object, Text, Text, DoubleWritable>{

    private Text orderId = new Text();
    private DoubleWritable weight = new DoubleWritable();
```

```
  public void map(Object key, Text value, Context context
                  ) throws IOException, InterruptedException {
    String[] fields = value.toString().split(",");
    if(fields[13].equals("Weight")){
        return;
    }
    orderId.set(fields[0]);
    weight.set(Double.parseDouble(fields[13]));
    context.write(orderId, weight);
  }
}

public static class WeightReducer
     extends Reducer<Text,DoubleWritable,Text,DoubleWritable> {
  private DoubleWritable result = new DoubleWritable();

  public void reduce(Text key, Iterable<DoubleWritable> values,
                     Context context
                     ) throws IOException, InterruptedException {
    int sum = 0;
    for (DoubleWritable val : values) {
      sum += val.get();
    }
    if (sum < 100) {
      result.set(sum);
      context.write(key, result);
    }
  }
}

public static void run() throws Exception {
```

```
    //Configuration conf = new Configuration();
    Job job = new Job();
    job.setJarByClass(OrderWeightLessThan100.class);
    job.setMapperClass(OrderMapper.class);
    job.setCombinerClass(WeightReducer.class);
    job.setReducerClass(WeightReducer.class);
    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(DoubleWritable.class);
    FileInputFormat.addInputPath(job, new Path("/sc"));
    HdfsFileDeleter.deleteFile(job, "/bel100");
    FileOutputFormat.setOutputPath(job, new Path("/bel100"));
    System.exit(job.waitForCompletion(true) ? 0 : 1);
  }
}
```

Objective 2

Mapper and reducer Function:

```
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import java.io.IOException;

public class DelayedOrders {
```

```
  public static class OrderMapper
       extends Mapper<Object, Text, Text, IntWritable>{

    private Text customer = new Text();
    private IntWritable lateDays = new IntWritable();

    public void map(Object key, Text value, Context context
                    ) throws IOException, InterruptedException {
      String[] fields = value.toString().split(",");
      if(fields[7].equals("Ship Late Day count")){
          return;
      }


customer.set(fields[8]);
      lateDays.set(Integer.parseInt(fields[7]));
      context.write(customer, lateDays);
    }
  }

  public static class LateDaysReducer
       extends Reducer<Text,IntWritable,Text,IntWritable> {
    private IntWritable totalLateDays = new IntWritable();

    public void reduce(Text key, Iterable<IntWritable> values,
                       Context context
                       ) throws IOException, InterruptedException {
      int sum = 0;
      int count = 0;
      for (IntWritable val : values) {
        sum += val.get();
```

```
      count++;
     }
     int averageLateDays = (int)Math.ceil((double)sum/count);
     totalLateDays.set(averageLateDays);
     context.write(key, totalLateDays);
    }
  }

  public static void run() throws Exception {
    //Configuration conf = new Configuration();
    Job job = new Job();
    job.setJarByClass(DelayedOrders.class);
    job.setMapperClass(OrderMapper.class);
    job.setCombinerClass(LateDaysReducer.class);
    job.setReducerClass(LateDaysReducer.class);
    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(IntWritable.class);
    FileInputFormat.addInputPath(job, new Path("/sc"));
    HdfsFileDeleter.deleteFile(job, "/delord");
    FileOutputFormat.setOutputPath(job, new Path("/delord"));
    System.exit(job.waitForCompletion(true) ? 0 : 1);
  }
}
```

Hfds File Deleter Code

```
import org.apache.hadoop.fs.FileSystem;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.mapreduce.Job;

public class HdfsFileDeleter {

  public static void deleteFile(Job job, String filePath) throws Exception {
    FileSystem fileSystem = FileSystem.get(job.getConfiguration());
    Path path = new Path(filePath);
    boolean deleted = fileSystem.delete(path, true);

    if (deleted) {
      System.out.println("Deleted file: " + filePath);
    } else {
      System.out.println("Failed to delete file: " + filePath);
    }
  }
}
```

OUTPUT

This dataset contains information on the flow of goods and materials through a supply chain network. It includes data on inventory levels, order fulfillment, transportation times, and other key metrics that are essential for managing logistics operations.

The dataset is organized into several tables, each representing a different aspect of the supply chain. The tables include:

Inventory: This table contains information on the inventory levels at each location in the supply chain, including warehouses, distribution centers, and retail stores. It includes data on the quantity of each product, the location of the inventory, and the date it was last updated.

Orders: This table contains information on customer orders, including the order date, order quantity, and destination. It also includes data on order fulfillment, such as the date the order was shipped and the tracking number.

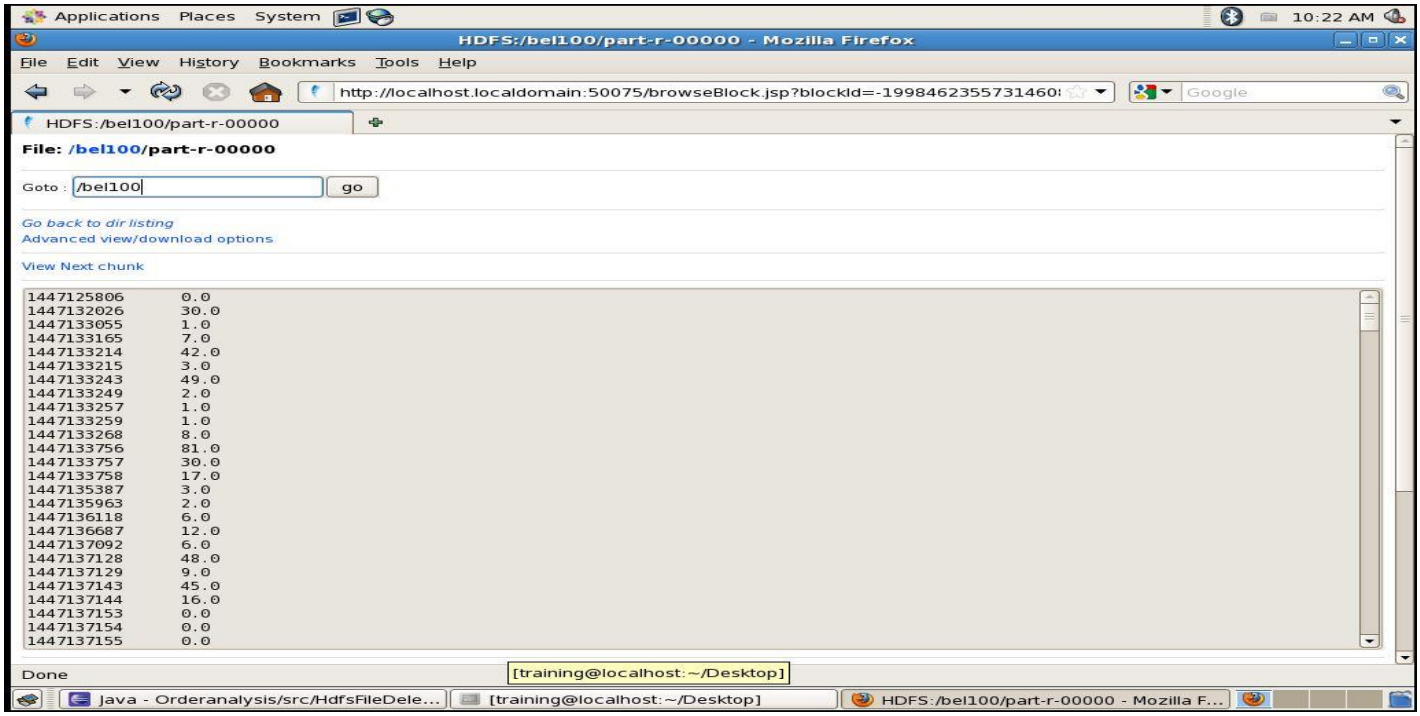
Transportation: This table contains information on the transportation of goods between locations in the supply chain. It includes data on the mode of transportation, transit time, and cost.

Suppliers: This table contains information on the suppliers of raw materials and finished goods. It includes data on the supplier name, location, lead time, and cost.

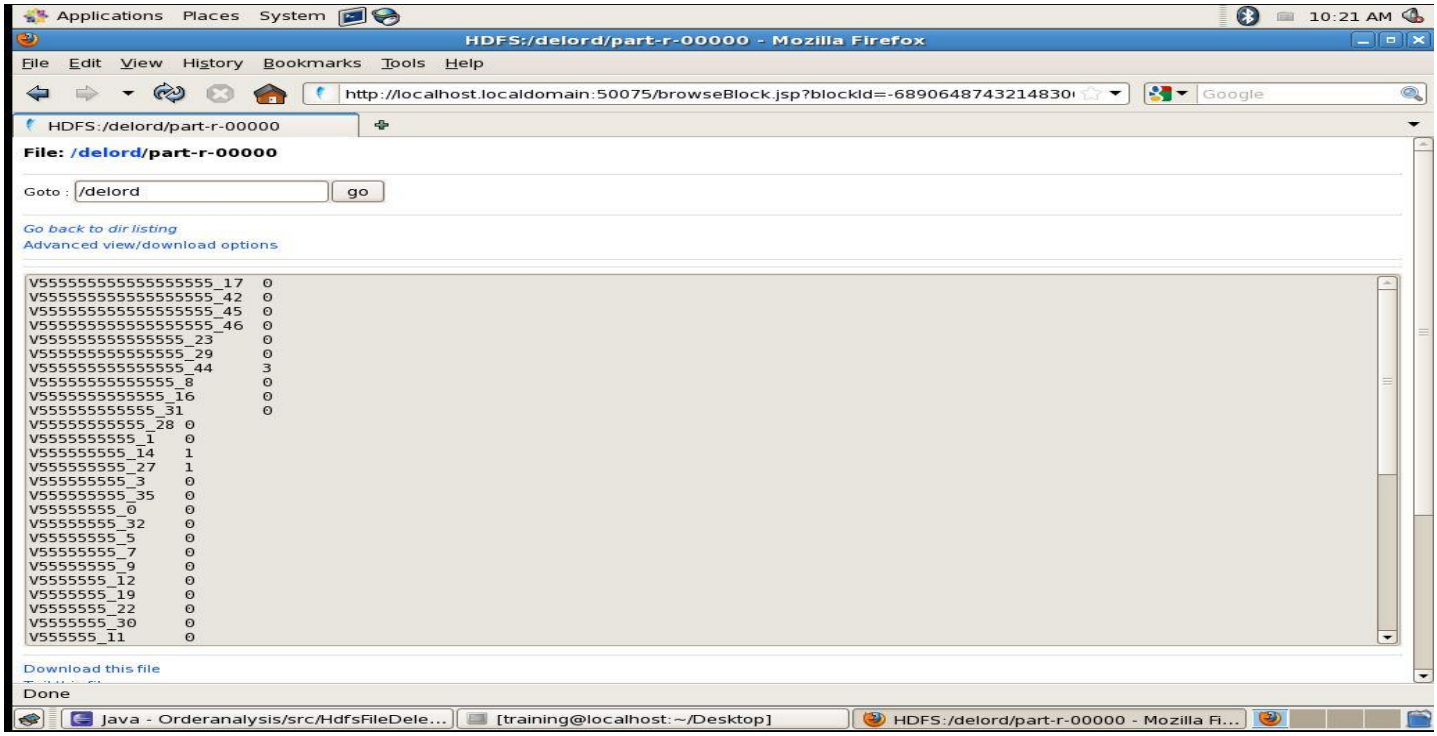
Overall, this dataset provides a comprehensive view of the supply chain network and can be used to optimize logistics operations, improve inventory management, and enhance customer satisfaction

Screenshots:

Objective 1



Objective 2:



OUTCOME OF PROJECT

- Display the order id with has the weight of less than 100
- To find the average days a customer has waited for the shipment to arrive.

CONCLUSION

In conclusion, having access to reliable and comprehensive datasets is crucial for effective supply chain and logistics management. With the increasing use of technology in these fields, there is an abundance of data available that can be used to optimize various aspects of the supply chain and logistics processes.

A well-organized and maintained dataset can provide valuable insights into factors such as inventory levels, transportation costs, delivery times, and demand patterns, enabling businesses to make informed decisions and improve their operations.

However, it is essential to ensure that the data is accurate, relevant, and up-to-date, and that appropriate data management practices are in place to maintain data integrity and security.

Furthermore, as the supply chain and logistics industries continue to evolve, it is likely that the types and volumes of data generated will continue to increase. Thus, it is essential to stay up-to-date with technological advancements and data analytics techniques to effectively leverage data and improve supply chain and logistics performance

REFERENCES

Ameri, F., & Patil, L. (2012). Digital manufacturing market: A semantic web-based framework for agile supply chain deployment. Journal of Intelligent Manufacturing, 23(5), 1817–1832. https://doi.org/10.1007/s10845-010-0495-z.

Ardito, L., Petruzzelli, A. M., Panniello, U., & Garavelli, A. C. (2019). Towards Industry 4.0: Mapping digital technologies for supply chain management-marketing integration. Business Process Management Journal.

Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. Computer Networks, 54(15), 2787–2805. https://doi.org/10.1016/j.comnet.2010.05.010.

Azuma, R. T. (2017). Making Augmented Reality a Reality. Imaging and Applied Optics 2017 (3D, AIO, COSI, IS, MATH, PcAOP), JTu1F.1. https://doi.org/10.1364/3D.2017.JTu1F.1.

Azzi, R., Chamoun, R. K., & Sokhn, M. (2019). The power of a blockchain-based supply chain. Computers & Industrial Engineering, 135, 582–592. https://doi.org/10.1016/j.cie.2019.06.042.

Babiceanu, R. F., & Seker, R. (2016). Big Data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook. Computers in Industry, 81, 128–137. https://doi.org/10.1016/j.compind.2016.02.004.

Backhaus, S. K. H., & Nadarajah, D. (2019). Investigating the relationship between industry 4.0 and productivity: A conceptual framework for Malaysian manufacturing firms. Procedía Computer Science, 161, 696–706.

Bai, C., Dallasega, P., Orzes, G., & Sarkis, J. (2020). Industry 4.0 technologies assessment: A sustainability perspective. International Journal of Production Economics, 229, 107776. https://doi.org/10.1016/j.ijpe.2020.107776.

Barholomae. (2018). Digital Transformation, International Competition and Specialization. 7.

Ben-Daya, M., Hassini, E., & Bahroun, Z. (2019). Internet of things and supply chain management: A literature review. International Journal of Production Research, 57(15–16), 4719–4742. https://doi.org/10.1080/00207543.2017.1402140.

Bhatti, Chandran, & Sundram. (2016). Supply chain practices and performance: The indirect effects of supply chain integration. https://www.emerald.com/insight/content/doi/10.1108/BIJ-03-2015-0023/full/html.