## INTRODUCTION TO DATA MANAGEMENT PROJECT REPORT

### (Project Semester August-December 2020)

***BIG FASHION SALES***

Submitted by

Parnasree Das

Registration No- 11802020

Programme- P132: B.Tech(Computer Science & Engineering) (2018) Section- KM069

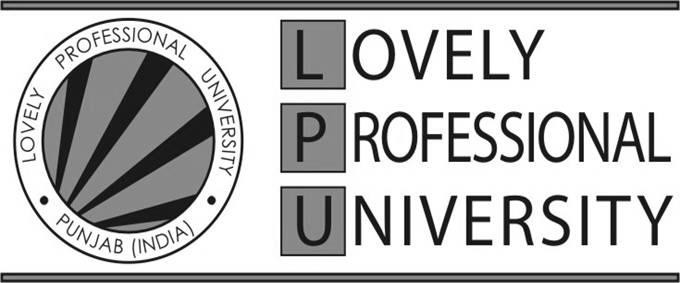
Course Code: INT-217

Under the Guidance of

**SAVLEEN with 18306 and Assistant Professor**

**Discipline of CSE/IT**

**Lovely School of technology and sciences Lovely Professional University, Phagwara**



**Declaration**

I **Parnasree Das** student of B.Tech under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 5th December 2020

Registration No: 11802020 Name of the student: Parnasree Das

**ACKNOWLEDGEMENT**

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I am highly indebted to *Miss Savleen* for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

I would like to express my gratitude towards my parents & member of Lovely Professional University for their kind co-operation and encouragement which help me in completion of this project.

My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities**.**

**Table of content Page number**

1. [Introduction 5](#_TOC_250004)
2. Objective of the analysis 6
3. [Source of data set… 6](#_TOC_250003)
4. [ETL Process… 7](#_TOC_250002)

[5. Analysis on data set……………………………………………………….](#_TOC_250001)

* 1. Objective 1… 8
  2. Objective 2… 9
  3. Objective 3… 11
  4. Objective 4… 13
  5. Objective 5… 14

1. List of analysis with result 17
2. [References… 18](#_TOC_250000)

#### INTRODUCTION

**About the Project:**

The final project of INT-217 (Introduction to Data Management) is the “**BIG FASHION SALES”**. This is the Project about to analyzing a dataset completely. For this project I have selected a dataset about fashions sales so we can perform different type of operation and formulas thereby making various types of visual representations, so it becomes easy to analyze.

This dataset is about sales data for two Australia wide fashion stores; Fashions

Direct and Next Look.

Table

Description automatically generatedThis dataset consists of 76,931 rows and 11 columns.

1. **OBJECTIVE OF THE ANALYSIS**
2. To analyze sales by chain
3. To analyze Sales by Category
4. To analyze sales by manager
5. To analyze sales percentage
6. To analyze Sales w.r.t States of Australia

## SOURCE OF DATA SET

The data for my study has been collected from [https://www.kaggle.com](https://www.kaggle.com/). It is a student friendly website with more than thousands of data to be practiced upon like the one I am doing with.

The data which I have selected is called TV Shows country wise. The link to my data set is:

https://www.kaggle.com/xv56ghbkli/fashion-sales

## ETL PROCESS

ETL is a process that extracts the data from the different source systems, then transforms the data (like applying calculations, concatenation, etc.) and finally loads the data into the Data Warehouse System. Its full form is Extract, Transform and load.

Steps to perform the ETL process are:

**Extraction**

Extraction is the first process where data from different sources like text file, XML file, Excel file, or various other sources are collected.

**Transformation**

Transformation is the second step of the ETL process, where all the collected data has been transformed into the same format. The format can be anything as per our requirement. In this step, a set of rules of functions are applied to the extracted data to convert it into a single standard format. It may involve the following tasks:

* **Filtering:** Only specific attributes are loading into the data warehouse.
* **Cleaning:** Filling up the null values with specific default values.
* **Joining:** Join the multiple attributes into the one.
* **Splitting:** Splitting the single attribute into multiple attributes.
* **Sorting:** Sort the tuples based on the attributes.

**Loading**

Loading is the final step of the ETL process. The big chunk of data is collected from various sources, transformed them, and finally loaded to the data warehouse. ETL is a process to extract the data from different source systems, transform the data, and load the data into the data warehouse. ETL process requires active input from various stakeholders including, developers, analysts, testers, top executive.

# ANALYSIS ON DATA SET

* + 1. **Objective 1:**  **to analyze sales by chain.**
    2. **Introduction:** In this given objective I am going to analyze sales by chain.
    3. **General Description: -** In the given objective we are going to create a pivot table and a pivot chart, then add slicer with respect to category and determined their success and failure.
    4. **Requirements:** For the following objective we do not need much only the followings are needed:
       - Windows 7 and above.
       - Tableau Prep Builder 2020.3 for cleaning
       - Microsoft Excel (Latest [2019] for power pivoting)
         * Pivot table
         * Pivot chart
         * Slicer
         * Timeline
         * Sorting

1. **Analysis results: -** In Mar 2017 the sales were the highest and that too in Fashions Direct and in every year sales of Next look is quite low as compared to Fashions Direct
2. **Visualisation:**

*Figure: 1*

Chart, line chart

Description automatically generated

**5.2 OBJECTIVE 2.** to analyze Sales by Category.

* + 1. **Introduction:** In this given objective I am going to to analyze Sales by Category.
    2. **General Description: -** In the given objective we are going to create a pivot table and a pivot chart, then add slicer with respect to category and determined their success and failure.
    3. **Requirements:** For the following objective we do not need much only the followings are needed:
* Windows 7 and above.
* Tableau Prep Builder 2020.3 for cleaning
* Microsoft Excel (Latest[2019] for power pivoting)
  + - * + Pivot table
        + Pivot chart
        + Slicer
        + Timeline
        + Sorting
    1. **Analysis results: -** Sales of Fashions Direct is high in all category and mens are the highest consumers of the product.
    2. *A picture containing chart

       Description automatically generated***Visulization**

***Figure: 2***

**5.3 OBJECTIVE 3. To analyze sales by manager.**

1. **Introduction:** In this given objective I am going to analyze sales by manager.
2. **General Description: -** In the given objective we are going to create a pivot table and a pivot chart, then add slicer with respect to year and determined their success and failure.
3. **Requirements:** For the following objective we do not need much only the followings are needed:
   * Windows 7 and above.
   * Tableau Prep Builder 2020.3 for cleaning
   * Microsoft Excel (Latest [2019] for power pivoting)
     + Pivot table
     + Pivot chart
     + Slicer
     + Timeline
     + Sorting
4. **Analysis results: -** Highest sales are by the Manager John Gardner. Sales in NSW is high as compared to other states of Australia. In states likes NSW, VIC, NT people prefer Fashions Direct while in States like SA, TAS people do prefer Next Look.
5. **Visualization:**

**A picture containing chart

Description automatically generated*Figure: 3***

* 1. **OBJECTIVE 4. to analyze sales percentage**

### **Introduction:** to analyse sales percentage

* + 1. **General Description: -** In the given objective we are going to create a pivot table and a pivot chart, then add slicer with respect to year and determined their success and failure.
    2. **Requirements:** For the following objective we do not need much only the followings are needed:
       - Windows 7 and above.
       - Tableau Prep Builder 2020.3 for cleaning
       - Microsoft Excel (Latest [2019] for power pivoting)
         * Pivot table
         * Pivot chart
         * Slicer
         * Timeline
         * Sorting
    3. **Analysis results: -** This table simply shows the sales percentage. As we can see Fashions Direct is having 70.74% of total sales.
    4. **Visualization:**

**Graphical user interface, text, application

Description automatically generated**

*Figure4*

* 1. **OBJECTIVE 5. Analyse Sales W.r.t States of Australia**

### **Introduction:** Analyse Sales W.r.t States of Australia

* + 1. **General Description: -** In the given objective we are going to create a pivot table and a pivot chart, then add slicer with respect to year and determined their success and failure.
    2. **Requirements:** For the following objective we do not need much only the followings are needed:
       - Windows 7 and above.
       - Tableau Prep Builder 2020.3 for cleaning
       - Microsoft Excel (Latest [2019] for power pivoting)
         * Pivot table
         * Pivot chart
         * Slicer
         * Timeline
         * Sorting
    3. **Analysis results: -** After analysis I found that Highest Sales are in NSW and i.e., 19895787.63 and Lowest Sales are in TAS and i.e., 1064824.81

14

* + 1. **Visualisation:**

**Text, table

Description automatically generatedA picture containing map

Description automatically generated**

*Figure: 5*

**List of Analysis with results**

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Analysis** | **Results** |
| **1.** | **to analyze sales by chain.** | In Mar 2017 the sales were the highest and that too in Fashions Direct and in every year sales of Next look is quite low as compared to Fashions Direct |
| **2.** | to analyze Sales by Category. | Sales of Fashions Direct is high in all category and mens are the highest consumers of the product. |
| **3.** | To analyze sales by manager | Highest sales are by the manager John Gardner. Sales in NSW is high as compared to other states of Australia. In states likes NSW, VIC, NT people prefer Fashions Direct while in States like SA, TAS people do prefer Next Look. |
| **4.** | **to analyze sales percentage** | After analysis I found that Highest Sales are in NSW and i.e., 19895787.63 and Lowest Sales are in TAS and i.e., 1064824.81 |
| **5.** | **Analyse Sales W.r.t States of Australia** | After analysis I found that Highest Sales are in NSW and i.e., 19895787.63 and Lowest Sales are in TAS and i.e., 1064824.81 |

1. **REFERENCES**
   1. MICROSOFT EXCEL 2016 BIBLE: THE COMPREHENSIVE TUTORIAL RESOURCE by JOHN WALKENBACH, WILEY
   2. FUNDAMENTALS OF BUSINESS ANALYTICS by R.N. PRASAD, SEEMA ACHARYA, WILEY
   3. WEBSITES: [https://www.kaggle.com](https://www.kaggle.com/)
   4. [**https://support.microsoft.com/en-us/excel?ui=en-us&rs=en-us&ad=us**](https://support.microsoft.com/en-us/excel?ui=en-us&rs=en-us&ad=us)
   5. [**https://edu.gcfglobal.org/en/excel2016/**](https://edu.gcfglobal.org/en/excel2016/)
   6. [**https://support.microsoft.com/en-us/excel?ui=en-us&rs=en-us&ad=us**](https://support.microsoft.com/en-us/excel?ui=en-us&rs=en-us&ad=us)
   7. [**https://excelexposure.com**](https://excelexposure.com/)
   8. [**https://chandoo.org/wp/excel-basics/**](https://chandoo.org/wp/excel-basics/)
   9. [**https://www.edx.org**](https://www.edx.org/)
   10. [**https://www.khanacademy.org/search?search\_again=1&page\_search\_que ry=microsoft+excel**](https://www.khanacademy.org/search?search_again=1&page_search_query=microsoft%2Bexcel)
   11. [**https://spreadsheetpage.com**](https://spreadsheetpage.com/)
   12. [**https://excel-bytes.com**](https://excel-bytes.com/)
   13. [**https://www.excelfunctions.net**](https://www.excelfunctions.net/)
2. **BIBLIOGRAPHY**
3. MICROSOFT EXCEL 2016 BIBLE: THE COMPREHENSIVE TUTORIAL RESOURCE by JOHN WALKENBACH, WILEY
4. FUNDAMENTALS OF BUSINESS ANALYTICS by R.N. PRASAD, SEEMA ACHARYA, WILEY
5. EXCEL 2016 FROM SCRATCH
6. EXCEL CHARTS FOR DUMMIES

**Projects:**

**Online Cab Booking:**

* **Skills** - Python, SQLite
* **Concepts Used** - Object-Oriented Programming, Database Handling , File handling, Exception Handling, GUI using Tkinter
* **Consists of -** Five pages that are welcome page which are linked to other three pages that are about us, new user and login page. Further Login page is linked with main page showing and fetching user information and gives the confirmation once the user is complete with the booking.

**Big Fashion Sale:**

* **Skills** - Excel
* **Concepts Used** - Data Cleaning, Data Analysis, Data Visualization
* **Consists of -** Eight pages such as file index, data , dashboard, results and analysis, etc. Performed various visualization techniques and came up with the insights such as analyzing Sales by chain, Sales by Category, Sales by the manager, Sales percentage, Sales w.r.t States of Australia and many more.

**Exploratory Data analysis of Supply Chain Shipment Price**

* **Skills** - Excel, Python
* **Concepts Used** - Data Cleaning, Data Analysis, Data Visualization
* **Consists of -** Python file showing various data cleaning techniques and then performing visualization and showing insights such as Frieght cost by shipment by country order , Frieght cost by Manufacturing Site, total freight cost for top 10 country and many more

**AWARDS AND RECOGNITIONS:**

* **Consolation Prize in Quiz Competition** | Ramakrishna Mission Ashram, Chandigarh | Jan 2014
* **1st Prize in International Informatics Olympiad** | Computer Literacy Foundation, New Delhi | 2008
* **10th Position in International Informatics Olympiad** | Computer Literacy Foundation, New Delhi |2009
* **Quiz Competition** | Shivalik Public School, Chandigarh |2014

**Academic CERTIFICATION:**

* **Data Structures And Algorithms Self Paced Course** | Geeks for Geeks | Jun 2020
* **Microlearning Course in Data Science** | Board Infinity | ID- BI20A2905038 | Jun 2020
* **Crash Course on Python** | Google | link- https://www.coursera.org/account/accomplishments/certificate/WWCPYHS2VYUQ | Grade Achieved: 93%| Apr2020
* **Google Cloud Platform Fundamentals: Core Infrastructure** | Google Cloud |link- https://www.coursera.org/account/accomplishments/certificate/6LNEADTRCHNE | Grade Achieved: 93% | Apr2020

**CONFERENCES AND WORKSHOPS:**

* MLH Local Hack Day | Organized by: Major League Hacking | Apr 2020
* Positive Psychology | Organized by The University of North Carolina at Chapel Hill | Apr 2020
* IOT, An Internet of Things Introductory Workshop |Organized by: Rayat-Bahra University |2017
* Participated in Foundation Day celebrations | Organized by: Indian Institute of Science Education and Research (IISER) |2016
* Teenovators |national level championship for innovation | Organized by: Manipal University |2014