# Global Sales Data Analytics

## A PROJECT REPORT

Submitted by

Jayachandran P

Jagadeesh S

Kalanjiya Vishnu J

Indrajith M

**TEAM ID**:**PNT2022TMID04402**

| SNO | TITTLE |
|---|---|
| **1** | **INTRODUCTION** |
| 1.1 | Project Overview |
| 1.2 | Purpose |
| **2** | **LITERATURE SURVEY** |
| 2.1 | Existing problem |
| 2.2 | References |
| 2.3 | Problem Statement Definition |
| **3** | **IDEATION &PROPOSED SOLUTION** |
| 3.1 | Empathy Map Canvas |
| 3 .2 | Ideation & Brainstorming |
| 3.3 | Proposed Solution |
| 3.4 | Problem Solution Fit |
| **4** | **REQUIREMENT ANALYSIS** |
| 4.1 | Functional requirements |
| 4.2 | Non-Functional requirements |
| **5** | **PROJECT DESIGN** |
| 5.1 | Data Flow Diagrams |
| 5.2 | Solution &Technical Architecture |
| 5.3 | User Stories |
| **6** | **PROJECT PLANNING & SCHEDULING** |
| 6.1 | Sprint Planning & Estimation |
| 6.2 | Sprint Delivery Schedule |
| 6.3 | Reports from JIRA |
| **7** | **CODING & SOLUTIONING** |
| 7.1 | Feature 1 |
| 7.2 | Feature 2 |
| 7.3 | Database Schema |
| **8** | **TESTING** |
| 8.1 | Test Cases |
| 8.2 | User Acceptance Testing |
| **9** | **RESULTS** |
| 9.1 | Performance Metrics |
| **10** | **ADVANTAGES & DISADVANTAGES** |
| **11** | **CONCLUSION** |
| **12** | **FUTURE SCOPE** |
| **13** | **APPENDIX** |

# 1.INTRODUCTION

If you want to achieve your sales goals month after month, then guesswork and intuition aren't your best friends. You need to perform a strategic sales analysis and get cold, hard data.You will gain an understanding of the data ecosystem and the fundamentals of data analysis, such as data gathering or data mining.

## 1.1 Project Overview:

The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support

## 1.2PURPOSE:

Regular sales data analysis provides an understanding of the products that your customers are buying and helps you dissect why they are behaving in a certain way. You can also find patterns in your lead conversions and drop offs.

Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions

Thousands of data points at your fingertips. Build, refine and analyse your audience in our intuitive platform. Monitor trends. Granular Global Analysis. 46 Countries. 17 Million Panelists. 40,000 Data Points. Create Bespoke Segments.

Sales analytics refers to the technology and processes used to gather sales data and gauge sales performance. Sales leaders use these metrics to set goals, improve internal processes, and forecast future sales and revenue more accurately.

## 2.LITERATURE SURVEY

### 2.1 Existing Problem:

1. Global sales process is way too long and don't have enough leads.
2. Leads are unqualified and wasting your effort on bad fit prospects.
3. Spending too much time on low-value task
4. The statement may include workflow bottlenecks,resources challenges or fundamental difficulties such as understanding a customer base
5. Identify the key sales metrics you need, such as win rate and average deal size
6. Use a tool (such as Pipe drive's CRM) to track this data as leads travel through your pipeline. Record this data in visual dashboards

### 2.2 REFERANCES:


1. Data mining with its role in marketing, sales support and customer identification data analysis [ Mohammed Bin Ali Al Atif, Ahmed H. Shakir, et al, 2022]
2. Impact of big data analytics on sales performance in pharmaceutical organizations: The role of customer relationship management capabilities [ Muhammad Shahbaz, Lili Zhai, et al, 2021]
3. Data Analysis and Visualization of Sales Dataset using Power BI [ Ms. Sarika Singh, Ms. Lavina Jadhav, 2022]
4. Survey on Growth of Business using Data Analytics for Business Intelligence in RealTime world [ Madamanchi Brahmaiah, Talluri Sreekrishna, 2021]

## 2.3 Problem Statement definition:

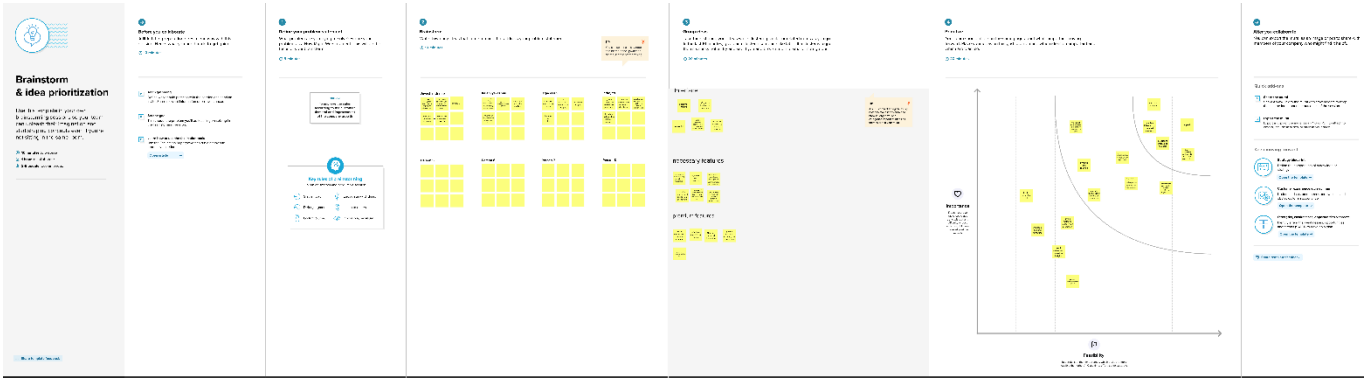
Customer Problem Statement :

- These days, online shopping is essential.
- So, try to comprehend a few things, such as the Global Super Store's Customer Analysis and Product Analysis.
- It's critical that the sales and marketing teams review their plans and effectiveness in order to enhance both.
- One method to gauge analysis of sales performance.
- The term "sales analytics" denotes the use of technology to gather and analyse sales data using data to generate practical knowledge.

## 3.IDEATION & PROPOSED SOLUTION

### 3.1 Empathy Map Canvas

# 3.2 Ideation & Brainstorming



## 3.3 Proposed Solution:

Project team shall fill the following information in proposed solution template.

| S.No. | Parameter | Description |
|---|---|---|
| 1. | Problem Statement (Problem to be solved) | Sales include all the actions involved in the product sale, consumer service and business service. For the sales and marketing team to review their performance data visualization techniques called sales analytics is used. In sales analytics, technology is used to collect and use the sales data to produce productive results and they are in turn used to identify and optimize the sales. Various attributes are used to plan an efficient sales model which will benefit both customer and business. |
| 2. | Idea / Solution description | Developing web application that would take all data and do analysis and give reports for visualization of data in a dashboard to identify trends for future analysis. |
| 3. | Novelty / Uniqueness | During the analysis, extraction of new features will be done. With that, more understanding can be made and we can come up with better decisions which will increase the salesperson's profit. |
| 4. | Social Impact / Customer Satisfaction | This sales data analytics improves the firms sales and future visions. and also Customer should know the available products and nearest location of the shops which gives the idea to customer for purchase. |
| 5. | Business Model (Revenue Model) | basic model: minimal analysis dashboard<br>subscription: reports on future prediction |
| 6. | Scalability of the Solution | the solution will be scalable even when the organisation productivity goes down |

## 3.4 Problem solution fit:

**Project Title: Global sales data analysis** **Project Design Phase-I -Solution Fit** **Team ID:** PNT2022TMID04402

Define CS, fit into CC

### 1. CUSTOMER SEGMENT(S)

e-commerce or business organization owners who wants to know the insights of their business and

### 6. CUSTOMER

The data needs to be collected in an organized form.
Decisions need to be made by the customer in all the situations.
Need more data to analyze more insights.

### 5. AVAILABLE SOLUTIONS

To create a dashboard with auto generated analytical insights.

Analyzing the data manually several algorithms and predefined library functions

Explore AS, differentiate

Focus on J&P, tap into BE, understand

### 2. JOBS-TO-BE-DONE / PROBLEMS

Analyzing the data and identifying the trends in the sales.

Determining the input file structure.

### 9. PROBLEM ROOT CAUSE

RC

Business model,
Change in trend,
The urge to do the job and earn money.

### 7. BEHAVIOUR

BE

Collecting sales data and using office software to analyze it.

Un-intuitive way of analyzing data and lot of manual labour.

Focus on J&P, tap into BE, understand

IDENTIFY STRONG TR & EM

### 3. TRIGGERS

TR

Don't know how to improve their business.

Fear of not going through the flow, not knowing the current trends.

The spark that their heart belongs for.

### 4. EMOTIONS: BEFORE / AFTER

EM

Before: anxiety, dicision fatigue, laziness
After: clear mind ,peacefullness

### 10. YOUR SOLUTION

SL

1.Creating an interactive dashboard.
2.responsive design for every screed size
3.one time payment

### 8. CHANNELS of BEHAVIOUR

CH

**8.1 ONLINE**

Using third party or subscription based softwares to analyze the data.

**8.2 OFFLINE**

Using office or selfdeveloped softwares to analyse the data

# 4.Requirement analysis:

## 4.1 Functional requirements :

Following are the functional requirements of the proposed solution.

| FR No. | Functional Requirement (Epic) | Sub Requirement (Story / Sub-Task) |
|---|---|---|
| FR-1 | User Registration | Registration through Form<br>Registration through Gmail<br>Registration through LinkedIN |
| FR-2 | User Confirmation | Confirmation via Email<br>Confirmation via OTP |
| FR-3 | Choose the tool for visualization | IBM Cognos analytics is chosen |
| FR-4 | Data visualization | Required graph, charts are chosen for visualization |
| FR-5 | Prepare dashboards | Dashboards, story boards and reports are created in IBM Cognos analytics |
| FR-6 | Business Decisions | Recommendations are made according to data |

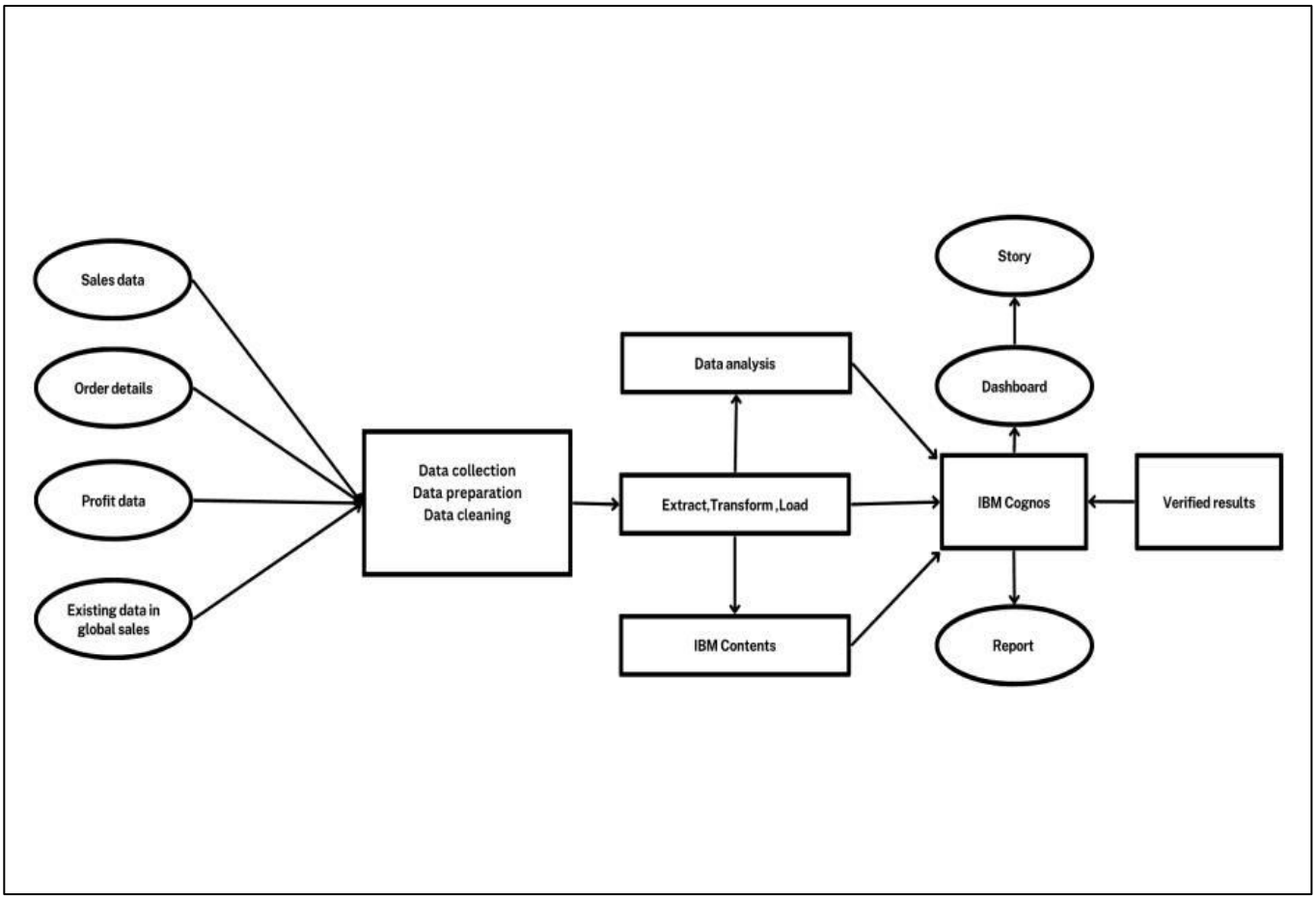
## 4.2 Non-Functional requirement:

Following are the non-functional requirements of the proposed solution.

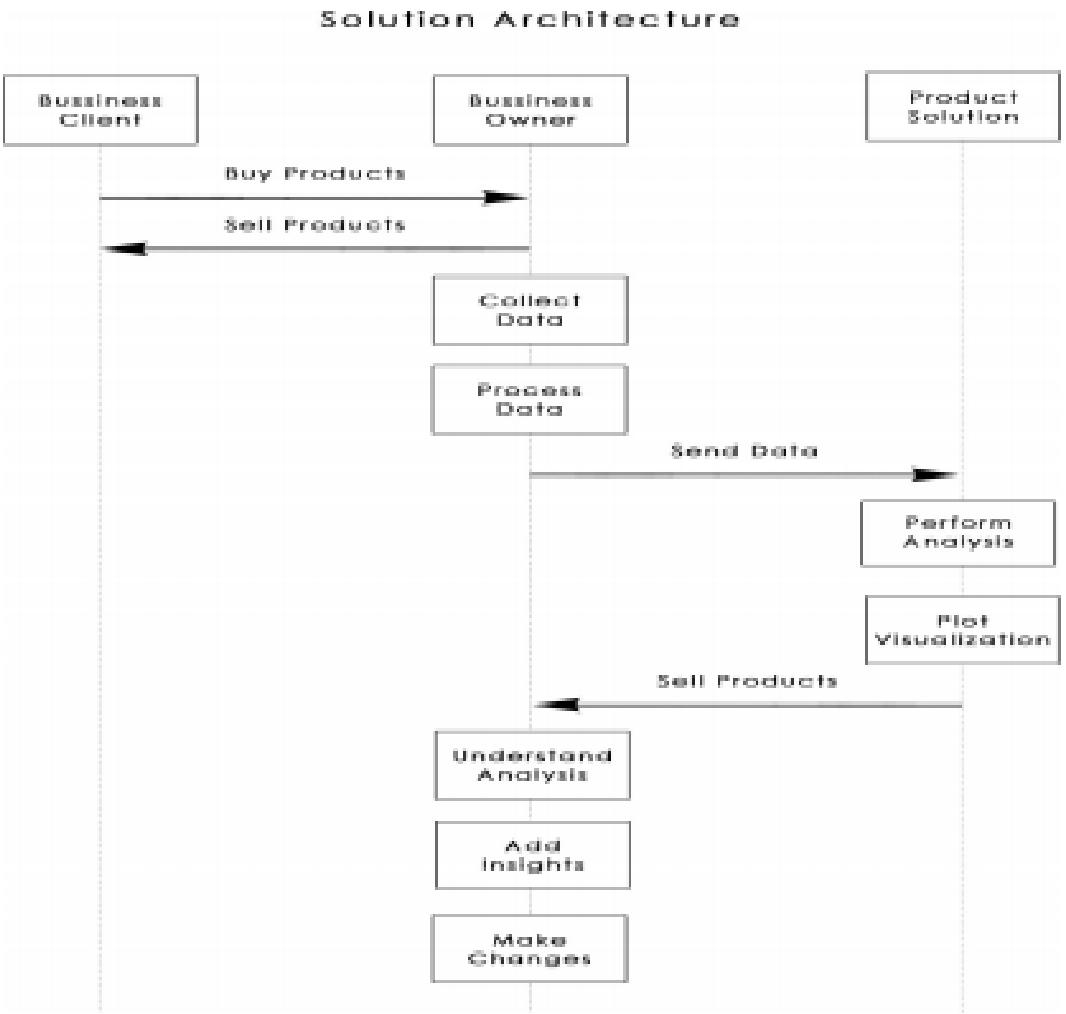
| FR No. | Non-Functional Requirement | Description |
|---|---|---|
| NFR-1 | **Usability** | It should be easier to understand the insights for the customers. |
| NFR-2 | **Security** | The data is protected from unauthorized access. |
| NFR-3 | **Reliability** | Connecting the data to the software and further process. |
| NFR-4 | **Performance** | The analysed information is recorded and updated. |
| NFR-5 | **Availability** | The tool is only available for the authorized persons to create, update, remove and the record customer information |
| NFR-6 | **Scalability** | Everyday activities are monitored for the growth of work.<br>Analytic tool should support even the size of data is increased. |

## 5.Project Design:

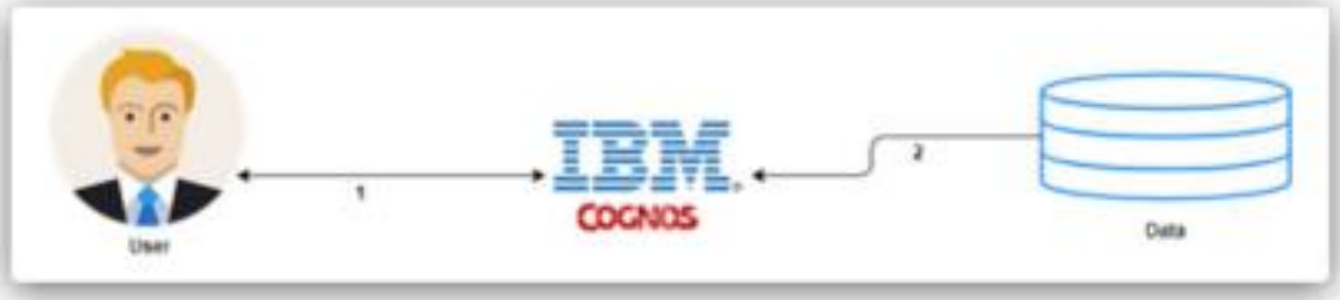
## 5.1.Data Flow Diagram:



## 5.2 Solution and Technical Architecture:




Technical Architecture :



## 6.Project Planning & Scheduling:

## 6.1 Sprint Planning & Estimation

| 7. 8. Sprint | Functional Requirement (Epic) | User Story Number | User Story / Task | Story Points |
|---|---|---|---|---|
| Sprint-1 | Collect the dataset | USN-1 | Download the dataset from Kaggle API | 2 |
| Sprint-1 | Understand the dataset | USN-2 | To understand the Data in dataset | 3 |
| Sprint-2 | Loading the dataset | USN-3 | Load the dataset in IBM cognos analytics | 3 |
| Sprint-2 | Preparation of dataset | USN-4 | Prepare the data with no null values | 4 |
| Sprint-2 | Performing calculations | USN-5 | Create new calculation for perfect visualization | 3 |

| Sprint | Functional Requirement (Epic) | User Story Number | User Story / Task | Story Points | Priority | Team Members |
|---|---|---|---|---|---|---|
| Sprint-1 | Collect the dataset | USN-1 | Download the dataset from Kaggle API | 2 | High | Indrajith, Kalanjiya Vishnu |
| Sprint-1 | Understand the dataset | USN-2 | To understand the Data in dataset | 3 | High | Indrajith, Kalanjiya Vishnu |
| Sprint-2 | Loading the dataset | USN-3 | Load the dataset in IBM cognos analytics | 3 | Low | Jagadeesh, Jayachandran, |
| Sprint-2 | Preparation of dataset | USN-4 | Prepare the data with no null values | 4 | Medium | Jagadeesh, Jayachandran, |
| Sprint-2 | Performing calculations | USN-5 | Create new calculation for perfect visualization | 3 | High | Jagadeesh, Indrajith, |

| Sprint | Functional Requirement (Epic) | User Story Number | User Story / Task | Story Points | Priority | Team Members |
|---|---|---|---|---|---|---|
| Sprint-3 | Creating visualization | USN-6 | Visualize the data for user to understand easily | 5 | Medium | Jayachandran, Indrajith, |
| Sprint-3 | Creating dashboard | USN-7 | To track, analyze and display data | 10 | Low | Jayachandran, Indrajith, |
| Sprint-4 | Report, Story and Final Delivery | USN-8 | Narratives that explain how fand why data changes over time, final delivery of the project | 20 | High | Jagadeesh, Kalanjiya vishnu |

### 3.1 Sprint Delivery Schedule:

| Sprint | Total Story Points | Duration | Sprint Start Date | Sprint End Date (Planned) | Story Points Completed (as on Planned End Date) | Sprint Release Date (Actual) |
|---|---|---|---|---|---|---|
| Sprint-1 | 5 | 6 Days | 24 Oct 2022 | 29 Oct 2022 | 5 | 29 Oct 2022 |
| Sprint-2 | 10 | 6 Days | 31 Oct 2022 | 05 Nov 2022 | 10 | 05 Nov 2022 |
| Sprint-3 | 15 | 6 Days | 07 Nov 2022 | 12 Nov 2022 | 15 | 12 Nov 2022 |
| Sprint-4 | 20 | 6 Days | 14 Nov 2022 | 19 Nov 2022 | 20 | 19 Nov 2022 |

**Velocity:**

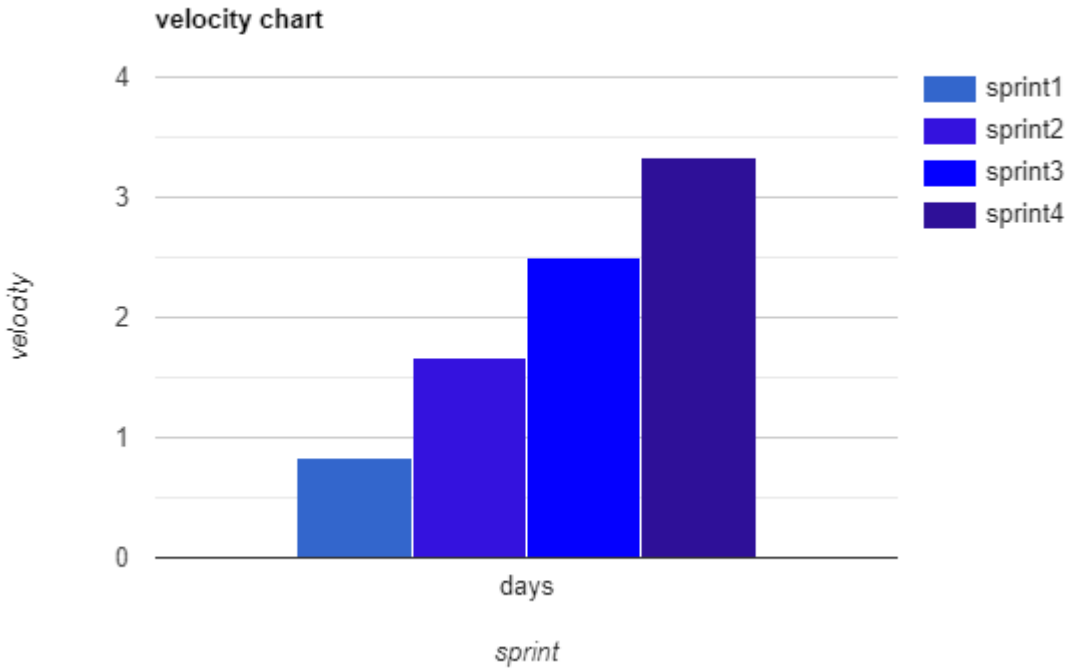
Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (story points per day)

$$AV = \frac{sprint\ duration}{velocity} = \frac{20}{10} = 2$$

AV=50/24=2.083

| Sprint | Total story points | Duration | Average velocity |
|---|---|---|---|
| Sprint-1 | 5 | 6 days | 5/6=0.833 |
| Sprint-2 | 10 | 6 days | 10/6=1.66 |
| Sprint-3 | 15 | 6 days | 15/6=2.5 |
| Sprint-4 | 20 | 6 days | 20/6=3.33 |

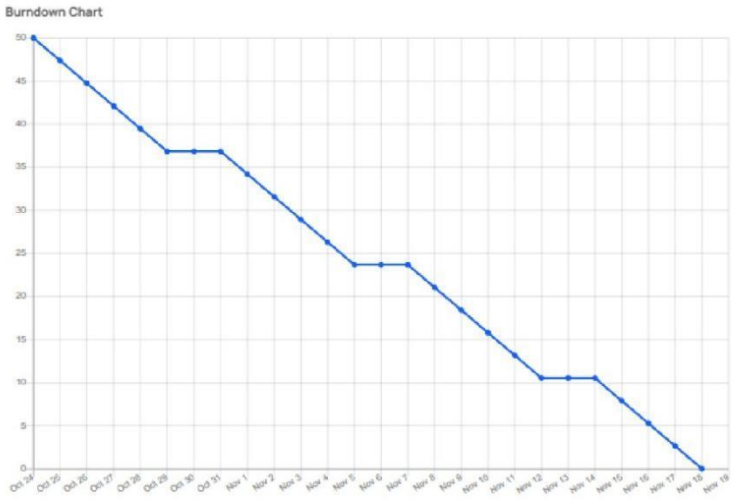
**Velocity chart:**



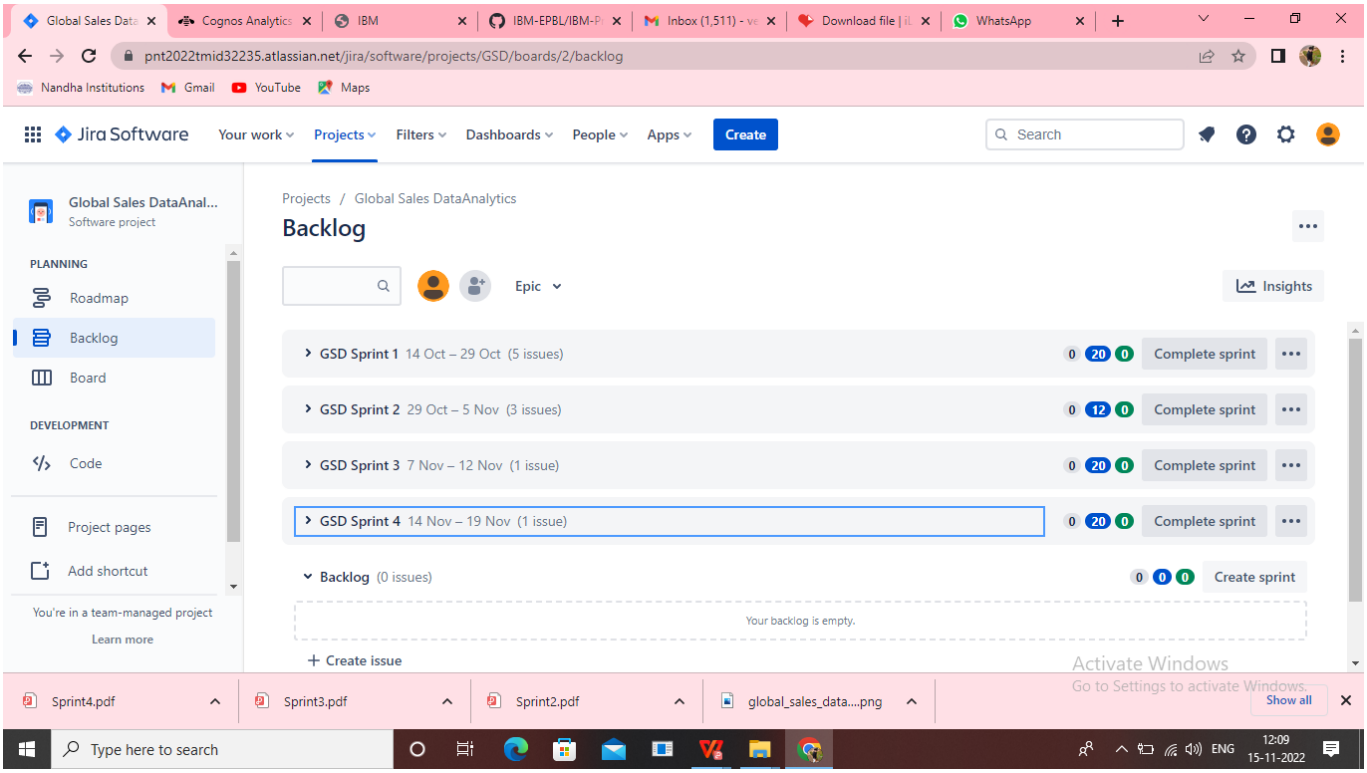
## Burndown Chart:

**Burndown Chart:**

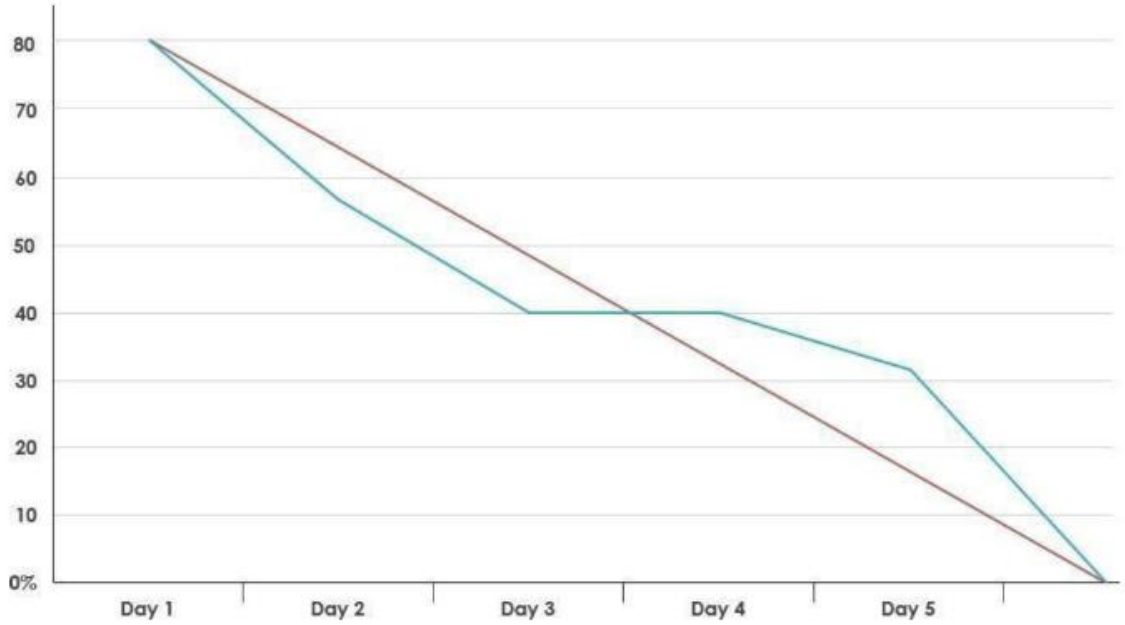
A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time.



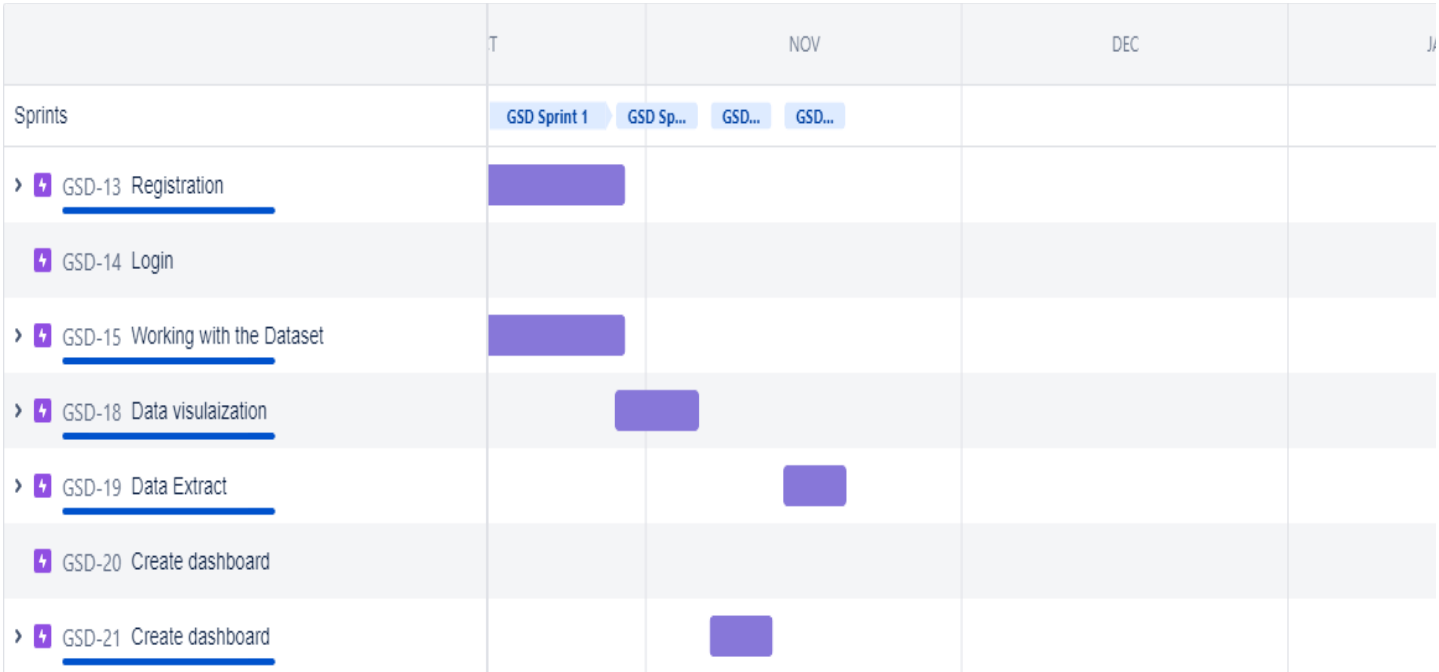
## 3.2 Reports from JIRA :



## Burndown chart :




## Road Map:



## 4.1 Feature 1

**Sales – Analysis:**

This is an analysis of the sales data with particular focus given to how promotions and advertising translate into sales, in terms of both units sold and sales dollars.

**Different types of Sales Analysis**

- Furniture company sales analysis HTML file
- Cereal Company Sales Analysis HTML file
- Financial Statement Analysis PDF file

**Analysis using R Shiny Dashboard**

- Furniture company sales Dashboard R Shiny app

**Steps for Cereal Company Sales Analysis**

1. Download the Raw Data
2. Analysis code R file
3. Final Analysis R file

**Steps for Furniture company sales analysis**

1. Download the Raw Data
2. Analysis code R file
3. Dashboard Code HTML file
4. Final Dashboard PDF file
5. Final Analysis HTML file

**fearture-1:**

**Step 1: Understand the Business**

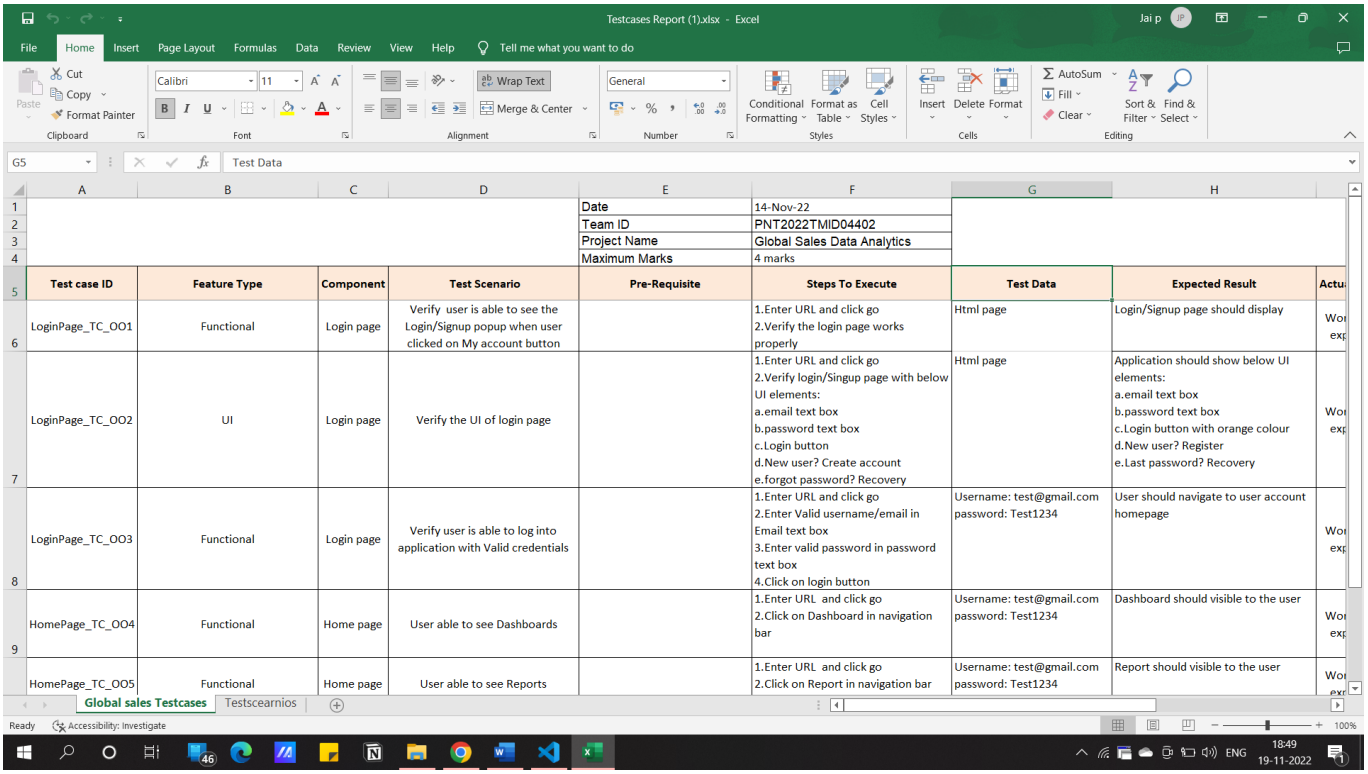
**Step 2: Get Your Data**

## Step 3: Explore and Clean Your Data

## Step 4: Enrich Your Datasets

# 5.Testing :

## 5.1 Test cases:



Testcases Report (1).xlsx - Excel

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Date | 14-Nov-22 | | | |
| | | | | Team ID | PNT2022TMID04402 | | | |
| | | | | Project Name | Global Sales Data Analytics | | | |
| | | | | Maximum Marks | 4 marks | | | |
| **Test case ID** | **Feature Type** | **Component** | **Test Scenario** | **Pre-Requisite** | **Steps To Execute** | **Test Data** | **Expected Result** | **Actu** |
| LoginPage_TC_OO1 | Functional | Login page | Verify user is able to see the Login/Signup popup when user clicked on My account button | | 1.Enter URL and click go<br>2.Verify the login page works properly | Html page | Login/Signup page should display | Wo<br>ex |
| LoginPage_TC_OO2 | UI | Login page | Verify the UI of login page | | 1.Enter URL and click go<br>2.Verify login/Singup page with below UI elements:<br>a.email text box<br>b.password text box<br>c.Login button<br>d.New user? Create account<br>e.forgot password? Recovery | Html page | Application should show below UI elements:<br>a.email text box<br>b.password text box<br>c.Login button with orange colour<br>d.New user? Register<br>e.Last password? Recovery | Wo<br>ex |
| LoginPage_TC_OO3 | Functional | Login page | Verify user is able to log into application with Valid credentials | | 1.Enter URL and click go<br>2.Enter Valid username/email in Email text box<br>3.Enter valid password in password text box<br>4.Click on login button | Username: test@gmail.com<br>password: Test1234 | User should navigate to user account homepage | Wo<br>ex |
| HomePage_TC_OO4 | Functional | Home page | User able to see Dashboards | | 1.Enter URL and click go<br>2.Click on Dashboard in navigation bar | Username: test@gmail.com<br>password: Test1234 | Dashboard should visible to the user | Wo<br>ex |
| HomePage_TC_OO5 | Functional | Home page | User able to see Reports | | 1.Enter URL and click go<br>2.Click on Report in navigation bar | Username: test@gmail.com<br>password: Test1234 | Report should visible to the user | Wo |

Global sales Testcases | Testscearnios

Ready | Accessibility: Investigate

ENG 18:49 19-11-2022

## 5.2 USER ACCEPTANCE TESTING

Copying and pasting screenshots of test results into Word or Excel is very time-consuming and prone to human error. Optimize your UAT testing with automated documentation, workflow and defect management. The right tool will help you with exploratory testing and be able to document tests using a recorder for playback as needed, accelerating the process and reducing the back-and-forth between the software development and testing teams.

## 6. RESULTS

### 6.1 PERFORMANCE Metrics:

The analysis covered the period from 2012 to 2015, with conversion to the Brazilian currency Real BRL (R$). Some results:

• The US was the country with the highest profit.

• The country that presented the biggest loss in sales was Turkey.

• There was greater demand for Superstore products to be shipped via the standard mode.

• The Technology Category presented better results in Profit and Sales.

• The Retail segment performed better for all the years evaluated.

## 7. ADVANTAGES

1. Cost efficiency
2. Receive full-scale services
3. Maximize presentation
4. Save time

## DISADVANTAGES

1. Risk of choosing the wrong provider
2. Lack of on-site support
3. Less control
4. Data security

## 8. CONCLUSION

By implementing this analytics solution, the company brought their competitive and sales data reporting in-house, cut costs and increased the accuracy of their reporting and analysis. As the company moves forward with this new solution, their sales reporting costs will most likely be reduced by 50 to 70%. They are now able to analyze raw data themselves, respond more quickly to changes in market trends and perform root cause analysis to determine those shifts in the market. By securing quicker access to their data with the new solution, the company was also able to reduce the risk associated with delayed responses to changes in their markets.With the new solution, the company can now process sales reports faster than the outsourced solution, reducing turnaround time between 50% to 60%. The reporting needs of the company have been streamlined, consolidating over 10reports into the centralized dashboard solution. The company's competitive analysis group is also able to more quickly respond to internal data requests given they have the ability to pull the information themselves. With this quicker response, the company is better able to react to changes in the market and predict opportunities for its sales force.The business also experienced an increase in the overall understanding of their sales data throughout the organization. The company now has great flexibility in the presentation of their sales and competitive data, while also being able to integrate sales data with other key data points for the organization.

## 9. FUTURE SCOPE

Sales analytics refers to the use of technology to collect and use sales data to derive actionable insights. It is used to identify, optimize, and forecast sales. It uses different metrics and KPIs to plan an efficient sales model that generates higher revenue for the business.

## 13.APPENDIX

## SOURCE CODE :

Index.tsx

```
import type { NextPage } from "next"
import { trpc } from "../utils/trpc"
import { IframeContainer } from "../components/IframeContainer"
import Link from "next/link"

const Home: NextPage = () => {
  const { data, isLoading } = trpc.useQuery(["graphs", { limit: 5 }])
  console.log("data", data)
  if (!data && isLoading) {
    return <div>Loading </div>
  }
  if (data) {
    return (
      <div className="mb-10">
        <Main />
        <ShowCase />
      </div>
    )
  }

  return <div className="p-8 bg-red-700"></div>
}

const ShowCase = () => {
  return (
    <div className="flex items-center justify-center h-max">
      <IframeContainer
        title="sales wise profit and discount"

link={`https://eu2.ca.analytics.ibm.com/bi/?perspective=explore&amp;pathRef=.my_folders%2Fdcasd%2FNew%2Bexploration&amp;closeWindowOnLastView=true&amp;ui_appbar=false&amp;ui_navbar=false&amp;shareMode=embedded&amp;subView=model000001848f3600a3_00000005`}
      />
    </div>
  )
```

```
}
const Main = () => {
  return (
    <main className="">
      <div className="relative px-6 lg:px-8">
        <div className="mx-auto max-w-3xl pt-20 pb-32 sm:pt-48 sm:pb-40">
          <div>
            <div className="hidden sm:mb-8 sm:flex sm:justify-center">
              <div className="relative overflow-hidden rounded-full py-1.5 px-4 text-
sm leading-6 ring-1 ring-gray-900/10 hover:ring-gray-900/20">
                <span className="text-gray-600">
                  Announcing our next round of funding.{" "}
                  <Link href="#">
                    <div className="font-semibold text-indigo-600">
                      <span className="absolute inset-0" aria-hidden="true" />
                      Read more <span aria-hidden="true">&rarr;</span>
                    </div>
                  </Link>
                </span>
              </div>
            </div>
            <div>
              <h1 className="text-4xl font-bold tracking-tight sm:text-center sm:text-
6xl">
                Analytics to enrich your online business
              </h1>
              <p className="mt-6 text-lg leading-8 text-gray-600 sm:text-center">
                Anim aute id magna aliqua ad ad non deserunt sunt. Qui irure qui
                lorem cupidatat commodo. Elit sunt amet fugiat veniam occaecat
                fugiat aliqua.
              </p>
              <div className="mt-8 flex gap-x-4 sm:justify-center">
                <Link href="/analytics">
                  <div className="inline-block rounded-lg bg-indigo-600 px-4 py-1.5
text-base font-semibold leading-7 text-white shadow-sm ring-1 ring-indigo-600
hover:bg-indigo-700 hover:ring-indigo-700">
                    Get started
                    <span className="text-indigo-200" aria-hidden="true">
                      &rarr;
                    </span>
```

```
            </div>
          </Link>
          <Link href="/dashboard">
            <span className="inline-block rounded-lg px-4 py-1.5 text-base font-semibold leading-7 text-gray-900 ring-1 ring-gray-900/10 hover:ring-gray-900/20">
              Dashboard
              <span className="text-gray-400" aria-hidden="true">
                &rarr;
              </span>
            </span>
          </Link>
        </div>
      </div>
      <div className="absolute inset-x-0 top-[calc(100%-13rem)] -z-10 transform-gpu overflow-hidden blur-3xl sm:top-[calc(100%-30rem)]">
        <svg
          className="relative left-[calc(50%+3rem)] h-[21.1875rem] max-w-none -translate-x-1/2 sm:left-[calc(50%+36rem)] sm:h-[42.375rem]"
          viewBox="0 0 1155 678"
          fill="none"
          xmlns="http://www.w3.org/2000/svg"
        >
          <path
            fill="url(#ecb5b0c9-546c-4772-8c71-4d3f06d544bc)"
            fillOpacity=".3"
            d="M317.219 518.975L203.852 678 0 438.341l317.219 80.634 204.172-286.402c1.307 132.337 45.083 346.658 209.733 145.248C936.936 126.058 882.053-94.234 1031.02 41.331c119.18 108.451 130.68 295.337 121.53 375.223L855 299l21.173 362.054-558.954-142.079z"
          />
          <defs>
            <linearGradient
              id="ecb5b0c9-546c-4772-8c71-4d3f06d544bc"
              x1="1155.49"
              x2="-78.208"
              y1=".177"
              y2="474.645"
              gradientUnits="userSpaceOnUse"
            >
```

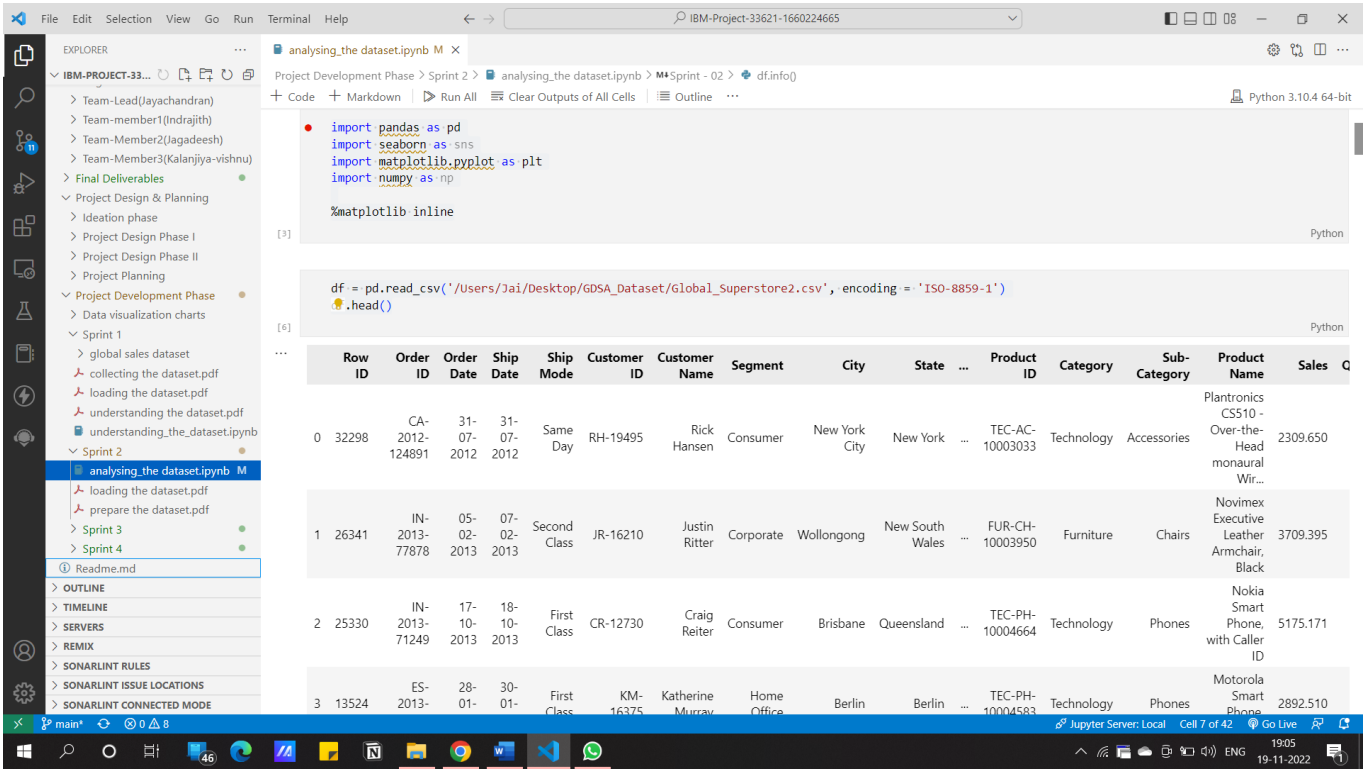
```
                <stop stopColor="#9089FC" />
                <stop offset={1} stopColor="#FF80B5" />
              </linearGradient>
            </defs>
          </svg>
        </div>
      </div>
    </div>
  </main>
 )
}

export default Home
```

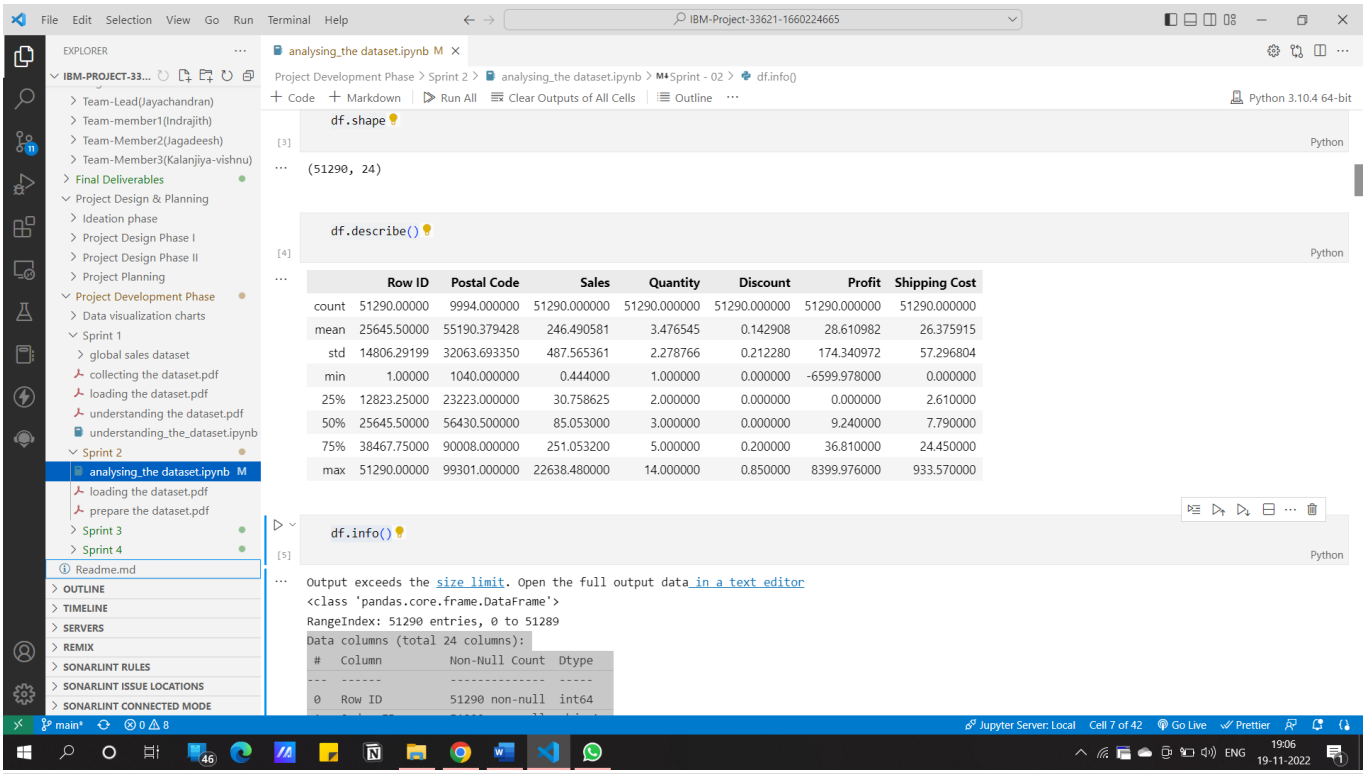
## Dashboard.tsx

```
import Banner from "../components/Banner"
import { IframeContainer } from "../components/IframeContainer"

export default function Dashboard() {
 return (
  <div className="flex flex-col gap-4">
   <Banner
    title="Sales analytics"
    subtitle="Sales analytics"
    tag="Learn More!"
   ></Banner>
   <div className="w-2/3 mx-auto mb-10 shadow-sm">
    <IframeContainer

link={`https://eu2.ca.analytics.ibm.com/bi/?perspective=dashboard&amp;pathRef=
.my_folders%2FDash%2BBoard%2BFinal&amp;closeWindowOnLastView=true
&amp;ui_appbar=false&amp;ui_navbar=false&amp;shareMode=embedded&amp;
action=view&amp;mode=dashboard&amp;subView=model000001848c07ed56_0
0000000`}
      width="100%"
      height="600"
     />
   </div>
  </div>
```

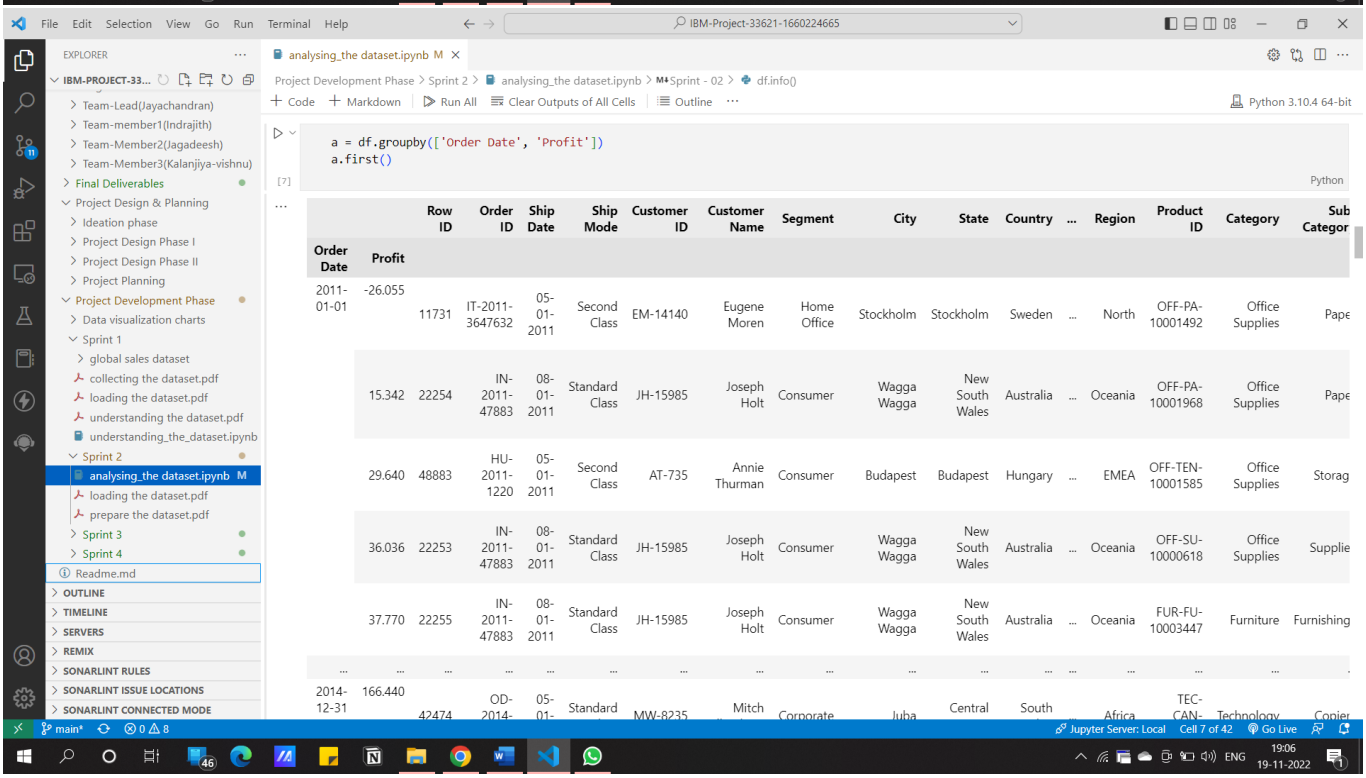
```
  )
}
```

**Most of the analysis is done in IBM-Cognos analytics tool but we did tried some analysis in python also .**

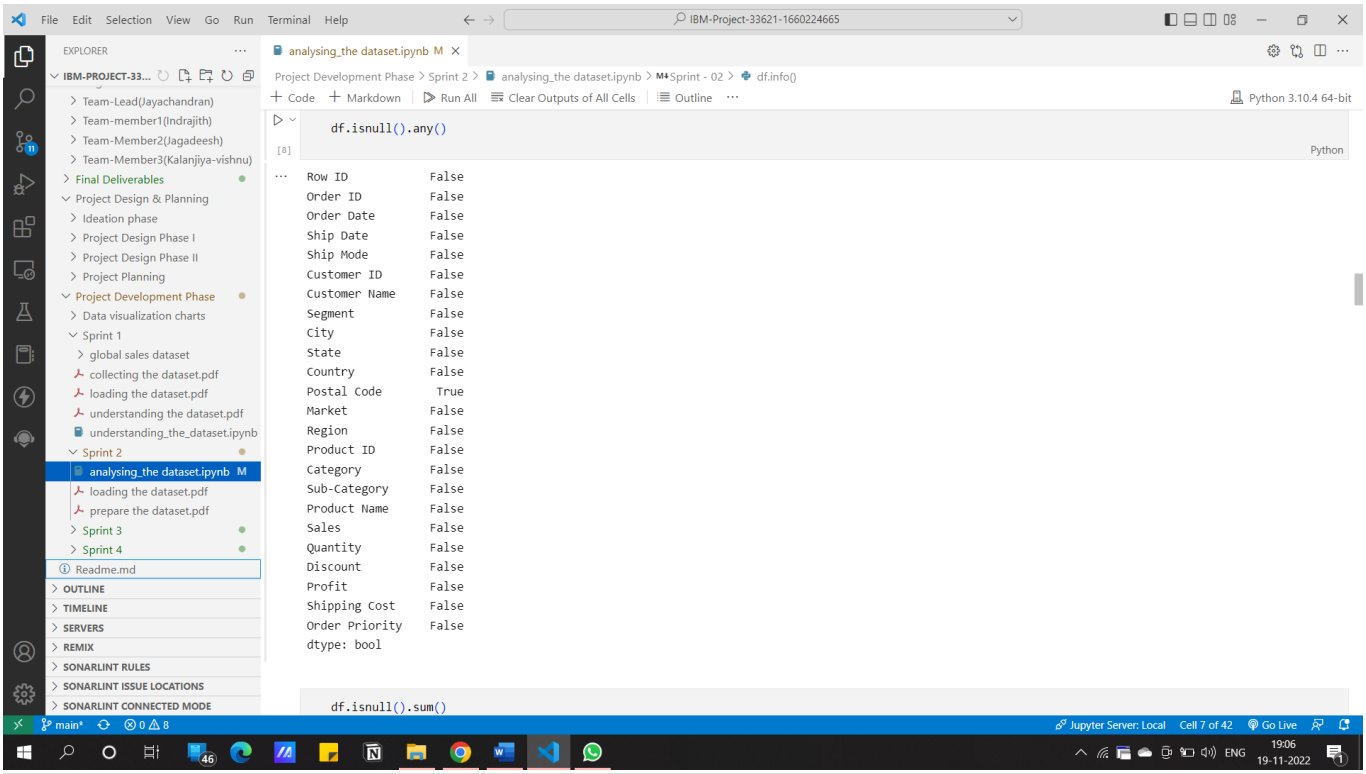


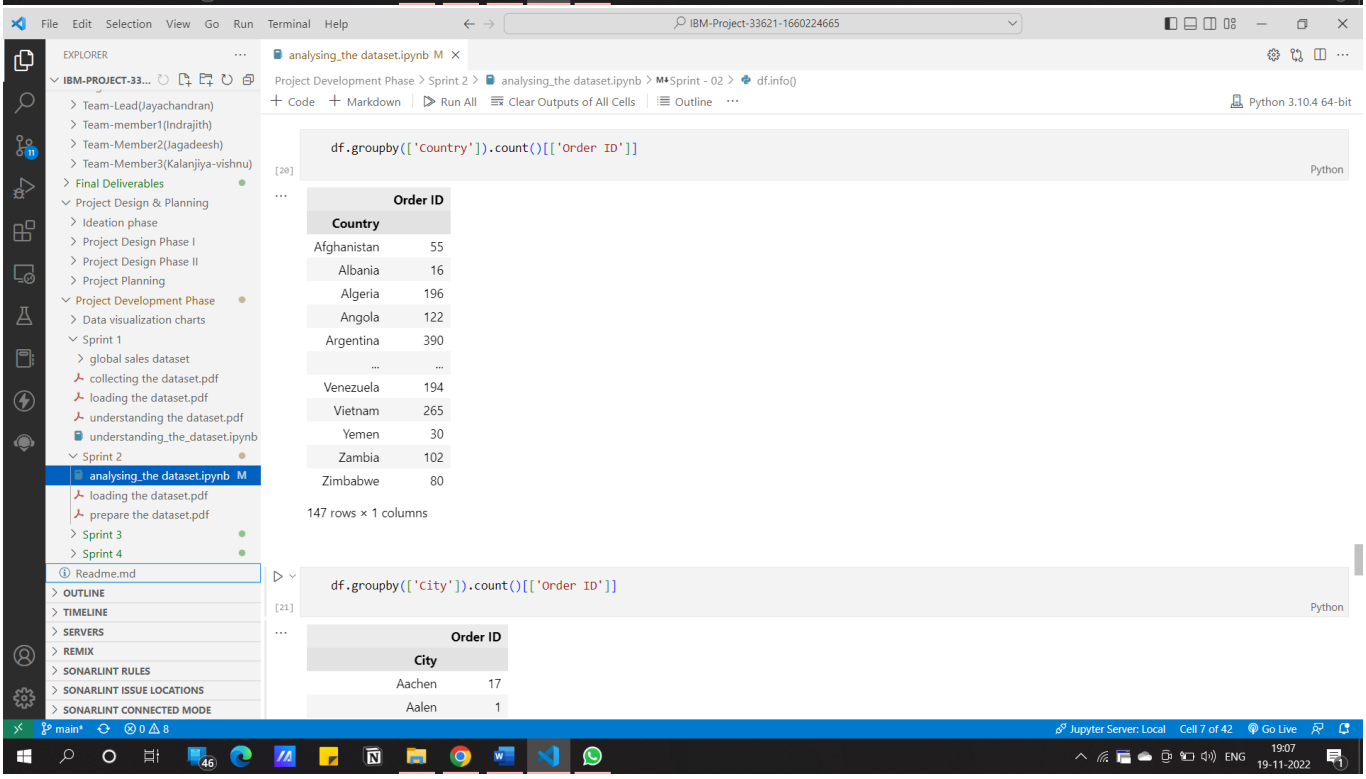


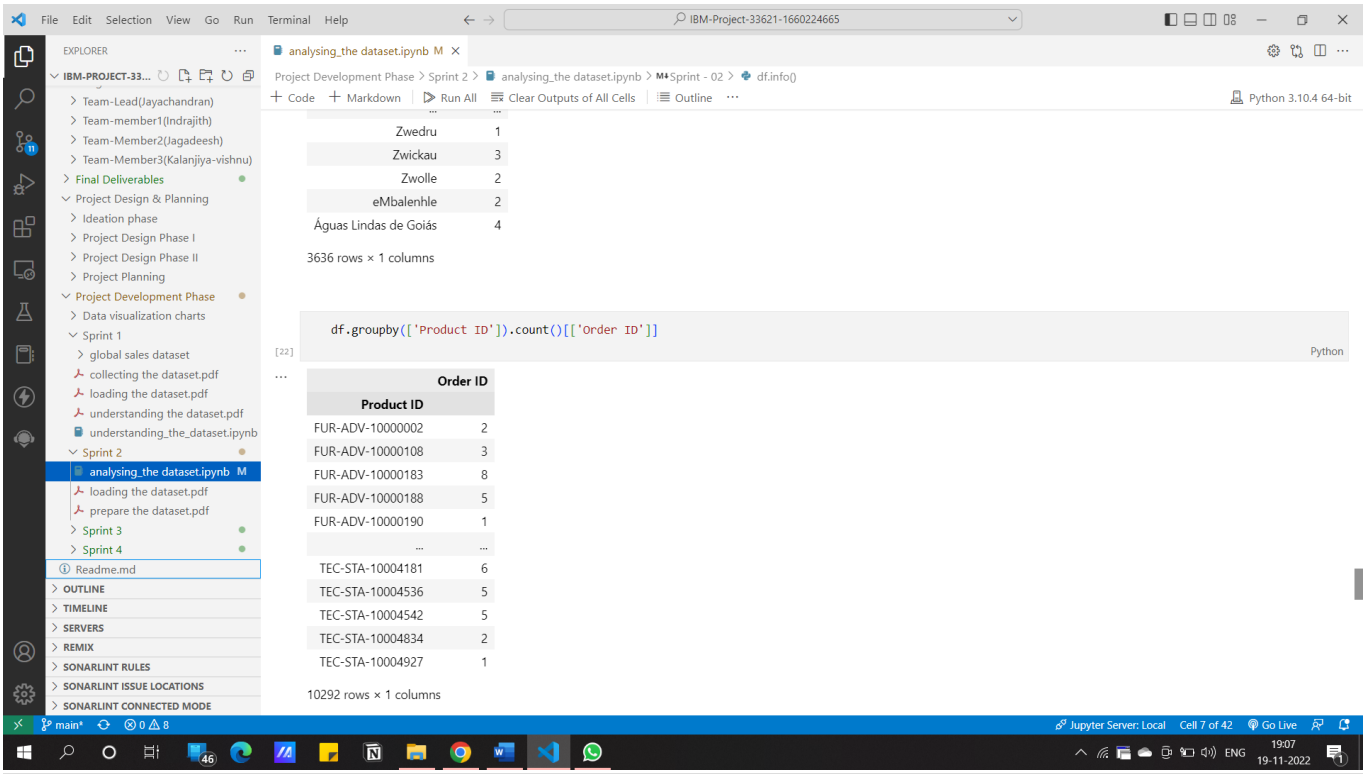

| | Row ID | Postal Code | Sales | Quantity | Discount | Profit | Shipping Cost |
|---|---|---|---|---|---|---|---|
| count | 51290.00000 | 9994.000000 | 51290.000000 | 51290.000000 | 51290.000000 | 51290.000000 | 51290.000000 |
| mean | 25645.50000 | 55190.379428 | 246.490581 | 3.476545 | 0.142908 | 28.610982 | 26.375915 |
| std | 14806.29199 | 32063.693350 | 487.565361 | 2.278766 | 0.212280 | 174.340972 | 57.296804 |
| min | 1.00000 | 1040.000000 | 0.444000 | 1.000000 | 0.000000 | -6599.978000 | 0.000000 |
| 25% | 12823.25000 | 23223.000000 | 30.758625 | 2.000000 | 0.000000 | 0.000000 | 2.610000 |
| 50% | 25645.50000 | 56430.500000 | 85.053000 | 3.000000 | 0.000000 | 9.240000 | 7.790000 |
| 75% | 38467.75000 | 90008.000000 | 251.053200 | 5.000000 | 0.200000 | 36.810000 | 24.450000 |
| max | 51290.00000 | 99301.000000 | 22638.480000 | 14.000000 | 0.850000 | 8399.976000 | 933.570000 |

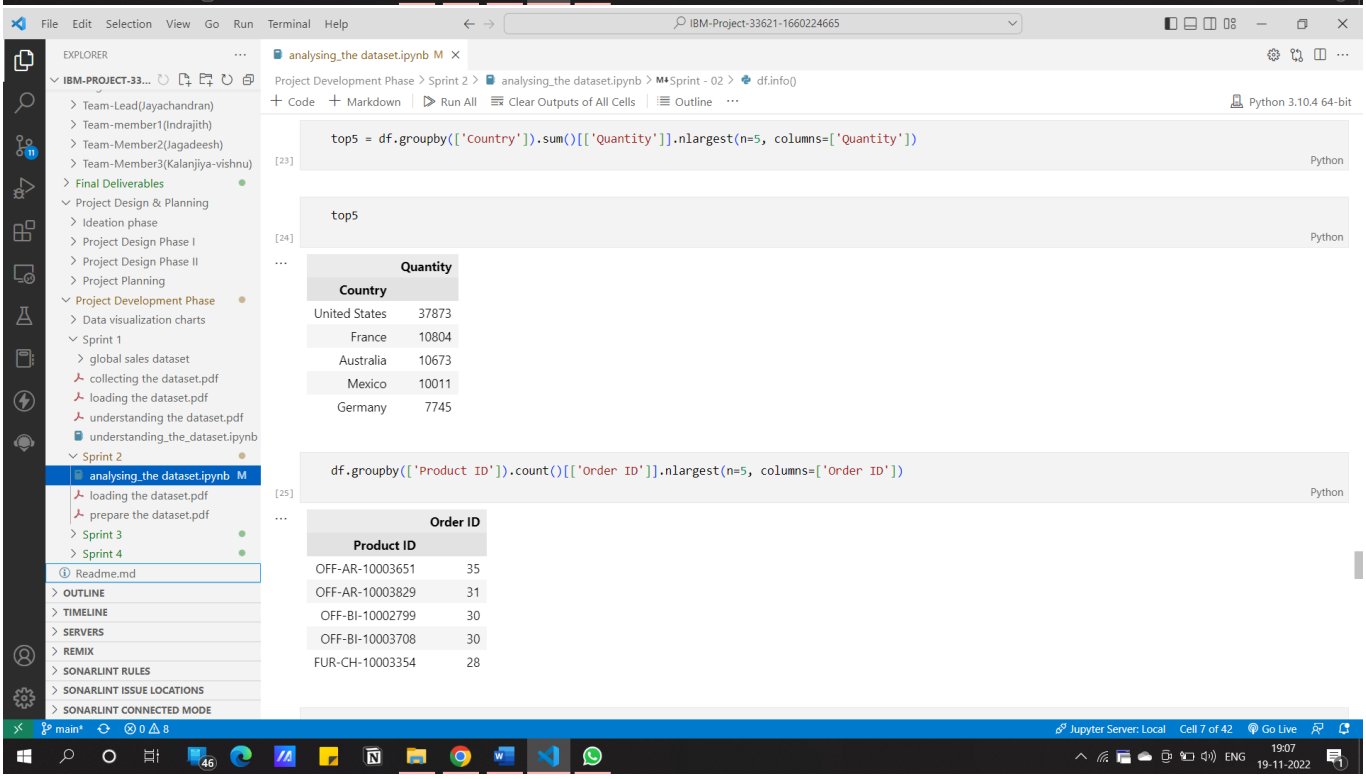



| Order Date | Profit | Row ID | Order ID | Ship Date | Ship Mode | Customer ID | Customer Name | Segment | City | State | Country | ... | Region | Product ID | Category | Sub Categor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011-01-01 | -26.055 | 11731 | IT-2011-3647632 | 05-01-2011 | Second Class | EM-14140 | Eugene Moren | Home Office | Stockholm | Stockholm | Sweden | ... | North | OFF-PA-10001492 | Office Supplies | Pape |
| | 15.342 | 22254 | IN-2011-47883 | 08-01-2011 | Standard Class | JH-15985 | Joseph Holt | Consumer | Wagga Wagga | New South Wales | Australia | ... | Oceania | OFF-PA-10001968 | Office Supplies | Pape |
| | 29.640 | 48883 | HU-2011-1220 | 05-01-2011 | Second Class | AT-735 | Annie Thurman | Consumer | Budapest | Budapest | Hungary | ... | EMEA | OFF-TEN-10001585 | Office Supplies | Storag |
| | 36.036 | 22253 | IN-2011-47883 | 08-01-2011 | Standard Class | JH-15985 | Joseph Holt | Consumer | Wagga Wagga | New South Wales | Australia | ... | Oceania | OFF-SU-10000618 | Office Supplies | Supplie |
| | 37.770 | 22255 | IN-2011-47883 | 08-01-2011 | Standard Class | JH-15985 | Joseph Holt | Consumer | Wagga Wagga | New South Wales | Australia | ... | Oceania | FUR-FU-10003447 | Furniture | Furnishing |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 2014-12-31 | 166.440 | 42474 | OD-2014- | 05-01- | Standard | MW-8235 | Mitch | Corporate | Juba | Central | South | ... | Africa | TEC-CAN- | Technology | Copier |


File Edit Selection View Go Run Terminal Help
IBM-Project-33621-1660224665
EXPLORER
analysing_the dataset.ipynb M
Project Development Phase > Sprint 2 > analysing_the dataset.ipynb > Sprint - 02 > df.info()
Code Markdown Run All Clear Outputs of All Cells Outline
Python 3.10.4 64-bit
df.isnull().any()
[8]
Python
Row ID False
Order ID False
Order Date False
Ship Date False
Ship Mode False
Customer ID False
Customer Name False
Segment False
City False
State False
Country False
Postal Code True
Market False
Region False
Product ID False
Category False
Sub-Category False
Product Name False
Sales False
Quantity False
Discount False
Profit False
Shipping Cost False
Order Priority False
dtype: bool
df.isnull().sum()
Jupyter Server: Local Cell 7 of 42 Go Live
19:06
19-11-2022


File Edit Selection View Go Run Terminal Help
IBM-Project-33621-1660224665
analysing_the dataset.ipynb M
Project Development Phase > Sprint 2 > analysing_the dataset.ipynb > Sprint - 02 > df.info()
Code Markdown Run All Clear Outputs of All Cells Outline
Python 3.10.4 64-bit
df.groupby(['Country']).count()[['Order ID']]
[20]
Python
Order ID
Country
Afghanistan 55
Albania 16
Algeria 196
Angola 122
Argentina 390
... ...
Venezuela 194
Vietnam 265
Yemen 30
Zambia 102
Zimbabwe 80
147 rows × 1 columns
df.groupby(['City']).count()[['Order ID']]
[21]
Python
Order ID
City
Aachen 17
Aalen 1
Jupyter Server: Local Cell 7 of 42 Go Live
19:07
19-11-2022


File Edit Selection View Go Run Terminal Help
IBM-Project-33621-1660224665
EXPLORER
analysing_the dataset.ipynb M
Project Development Phase > Sprint 2 > analysing_the dataset.ipynb > Sprint - 02 > df.info()
Code Markdown Run All Clear Outputs of All Cells Outline
Python 3.10.4 64-bit
Zwedru 1
Zwickau 3
Zwolle 2
eMbalenhle 2
Águas Lindas de Goiás 4
3636 rows × 1 columns
df.groupby(['Product ID']).count()[['Order ID']]
[22]
Python
Order ID
Product ID
FUR-ADV-10000002 2
FUR-ADV-10000108 3
FUR-ADV-10000183 8
FUR-ADV-10000188 5
FUR-ADV-10000190 1
TEC-STA-10004181 6
TEC-STA-10004536 5
TEC-STA-10004542 5
TEC-STA-10004834 2
TEC-STA-10004927 1
10292 rows × 1 columns
Team-Lead(Jayachandran)
Team-member1(Indrajith)
Team-Member2(Jagadeesh)
Team-Member3(Kalanjiya-vishnu)
Final Deliverables
Project Design & Planning
Ideation phase
Project Design Phase I
Project Design Phase II
Project Planning
Project Development Phase
Data visualization charts
Sprint 1
global sales dataset
collecting the dataset.pdf
loading the dataset.pdf
understanding the dataset.pdf
understanding_the_dataset.ipynb
Sprint 2
analysing_the dataset.ipynb M
loading the dataset.pdf
prepare the dataset.pdf
Sprint 3
Sprint 4
Readme.md
OUTLINE
TIMELINE
SERVERS
REMIX
SONARLINT RULES
SONARLINT ISSUE LOCATIONS
SONARLINT CONNECTED MODE
main*
Jupyter Server: Local
Cell 7 of 42
Go Live
ENG
19:07
19-11-2022


File Edit Selection View Go Run Terminal Help
IBM-Project-33621-1660224665
EXPLORER
analysing_the dataset.ipynb M
Project Development Phase > Sprint 2 > analysing_the dataset.ipynb > Sprint - 02 > df.info()
Code Markdown Run All Clear Outputs of All Cells Outline
Python 3.10.4 64-bit
top5 = df.groupby(['Country']).sum()[['Quantity']].nlargest(n=5, columns=['Quantity'])
[23]
Python
top5
[24]
Python
Quantity
Country
United States 37873
France 10804
Australia 10673
Mexico 10011
Germany 7745
df.groupby(['Product ID']).count()[['Order ID']].nlargest(n=5, columns=['Order ID'])
[25]
Python
Order ID
Product ID
OFF-AR-10003651 35
OFF-AR-10003829 31
OFF-BI-10002799 30
OFF-BI-10003708 30
FUR-CH-10003354 28
Team-Lead(Jayachandran)
Team-member1(Indrajith)
Team-Member2(Jagadeesh)
Team-Member3(Kalanjiya-vishnu)
Final Deliverables
Project Design & Planning
Ideation phase
Project Design Phase I
Project Design Phase II
Project Planning
Project Development Phase
Data visualization charts
Sprint 1
global sales dataset
collecting the dataset.pdf
loading the dataset.pdf
understanding the dataset.pdf
understanding_the_dataset.ipynb
Sprint 2
analysing_the dataset.ipynb M
loading the dataset.pdf
prepare the dataset.pdf
Sprint 3
Sprint 4
Readme.md
OUTLINE
TIMELINE
SERVERS
REMIX
SONARLINT RULES
SONARLINT ISSUE LOCATIONS
SONARLINT CONNECTED MODE
main*
Jupyter Server: Local
Cell 7 of 42
Go Live
ENG
19:07
19-11-2022

```
top5 = df.groupby(['Country']).sum()[['Quantity']].nlargest(n=5, columns=['Quantity'])
```

[26] Python

```
df2 = df.groupby(['Product Name']).sum()[['Profit']].nlargest(n=5, columns=['Profit'])
df2
```

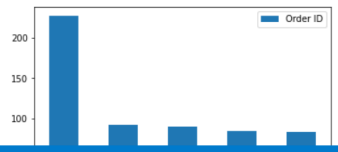
[27] Python

| Product Name | Profit |
|---|---|
| Canon imageCLASS 2200 Advanced Copier | 25199.9280 |
| Cisco Smart Phone, Full Size | 17238.5206 |
| Motorola Smart Phone, Full Size | 17027.1130 |
| Hoover Stove, Red | 11807.9690 |
| Sauder Classic Bookcase, Traditional | 10672.0730 |

TOP 5 PRODUCT BY TOTAL ORDER

```
df.groupby(['Product Name']).count()[['Order ID']].sort_values(by="Order ID",ascending=False).nlargest(n=5, columns=['Order ID']).plot.bar()
plt.show()
```

[28] Python

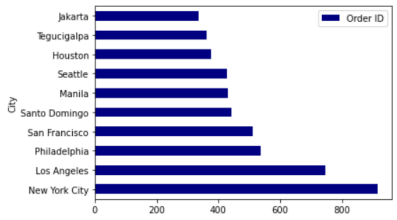



| Product Name | Order ID |
|---|---|
| Staples | 227 |
| Cardinal Index Tab, Clear | 92 |
| Eldon File Cart, Single Width | 90 |
| Rogers File Cart, Single Width | 84 |
| Ibico Index Tab, Clear | 83 |

TOP 10 CITY BY TOTAL ORDER

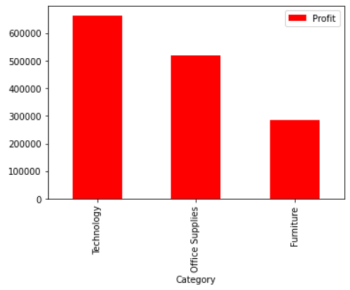
```
df.groupby(['City']).count()[['Order ID']].sort_values(by="Order ID",ascending=True).nlargest(n=10, columns=['Order ID']).plot.barh(color='navy')
plt.show()
```

[30] Python



```
df.groupby(['Category']).sum()[['Profit']].sort_values(by="Profit",ascending=False).nlargest(n=5, columns=['Profit']).plot.bar(color="Red")
plt.show()
```

[35] Python




TOTAL ORDER BY CATEGORY

```
df.groupby(['Category']).count()[['Order ID']].sort_values(by="Order ID",ascending=False).nlargest(n=5, columns=['Order ID']).plot.pie(subplots=True)
plt.show()
```

[36] Python



```
df.groupby(['Category']).sum()[['Profit']].sort_values(by="Profit",ascending=False).nlargest(n=5, columns=['Profit']).plot.pie(subplots=True)
plt.show()
```

[37] Python