Project Closure Document

Accelerating Analytics with Data bricks and AWS S3.

1. DASHBOARD USER MANUAL

1.1 Getting Started

This section briefs about the minimum system requirements required to access the dashboard, the procedure to install Power BI and the procedure to import the dashboard.

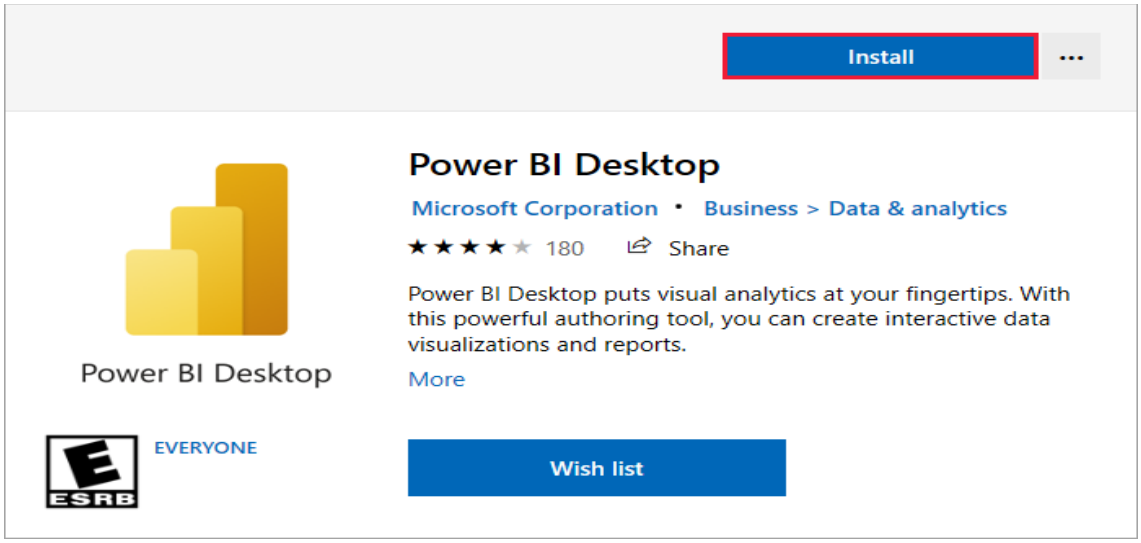
1.2 Minimum System Requirements

- Windows 8.1 or Windows Server 2012 R2 or later.
- .NET 4.6.2 or later.
- Microsoft Edge browser (Internet Explorer is no longer supported)
- Memory (RAM): At least 2 GB available, 4 GB or more recommended.
- Display: At least 1440x900 or 1600x900 (16:9) required. Lower resolutions such as 1024x768 or 1280x800 aren't supported because some controls (such as closing the startup screens) display beyond those resolutions.
- CPU: 1 gigahertz (GHz) 64-bit (x64) processor or better recommended.

1.3 Procedure to Install Power BI

1) Install as an app from the Microsoft Store

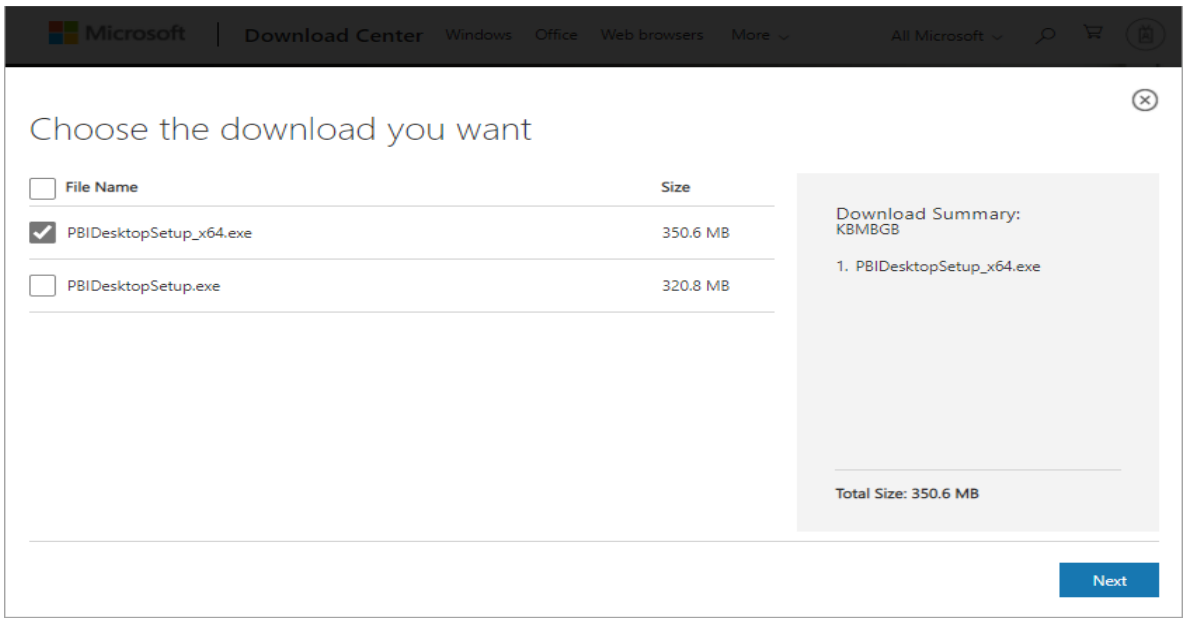
- Open a browser and go directly to the Power BI Desktop page of the Microsoft Store
- After you've landed on the Power BI Desktop page of the Microsoft Store, select install.



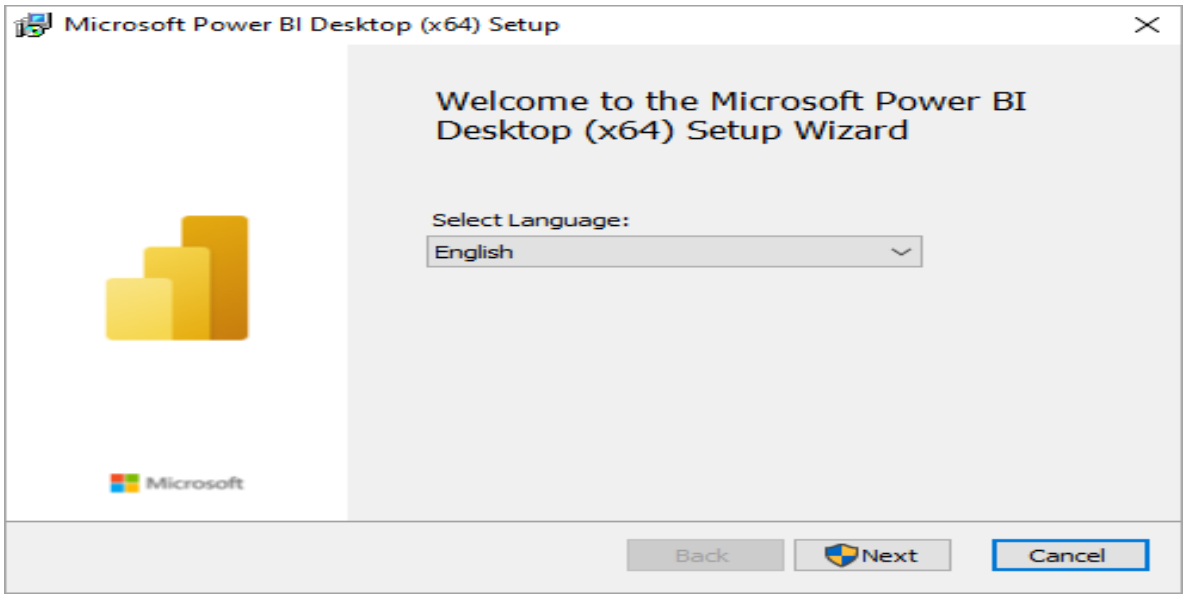
- Now Power BI Desktop will Installed in your device and will be readily accessible.

2) Install Power BI Desktop after download

- To download the Power BI Desktop executable from the Download Center, select Download from the Download Center page. Then specify the 32-bit or 64-bit installation file to download.



- After you launch the installation package, Power BI Desktop installs as an application and runs on your desktop.



1.4 The Store Manager Dashboard

1.4.1 About the dashboard

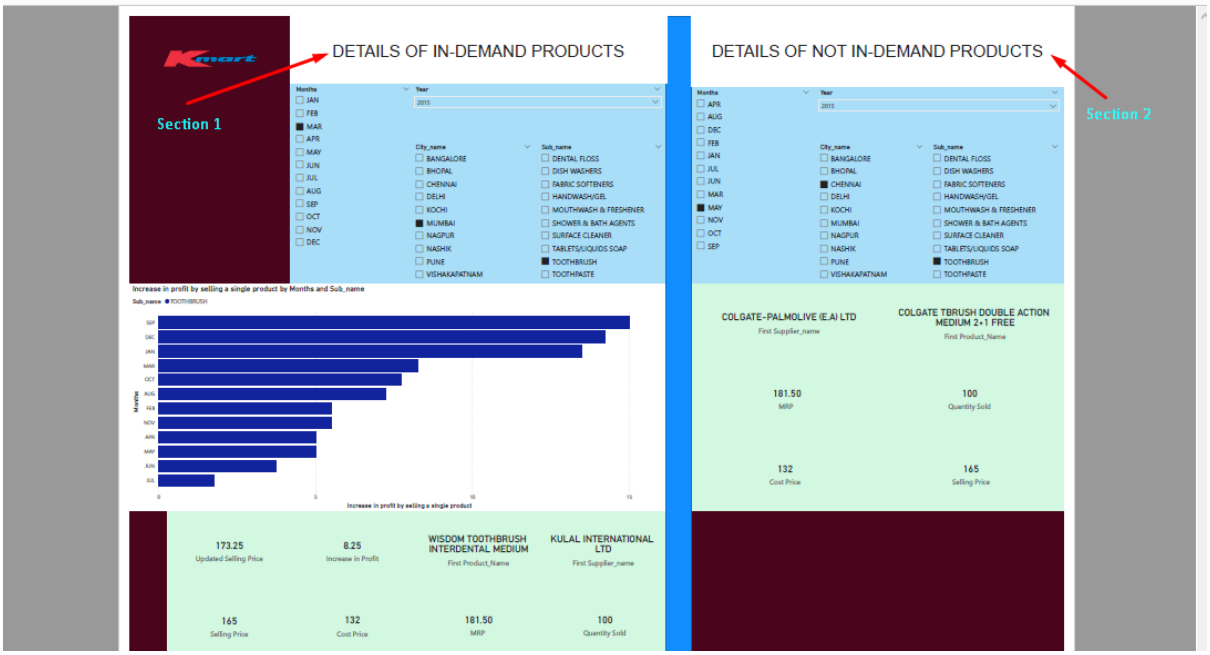
This dashboard is divided into two sections

- In-demand products: Contains the detail about the product with maximum sales in a particular product type in a particular month and in a particular year. And also the new selling price of the product along with the supplier of the product.

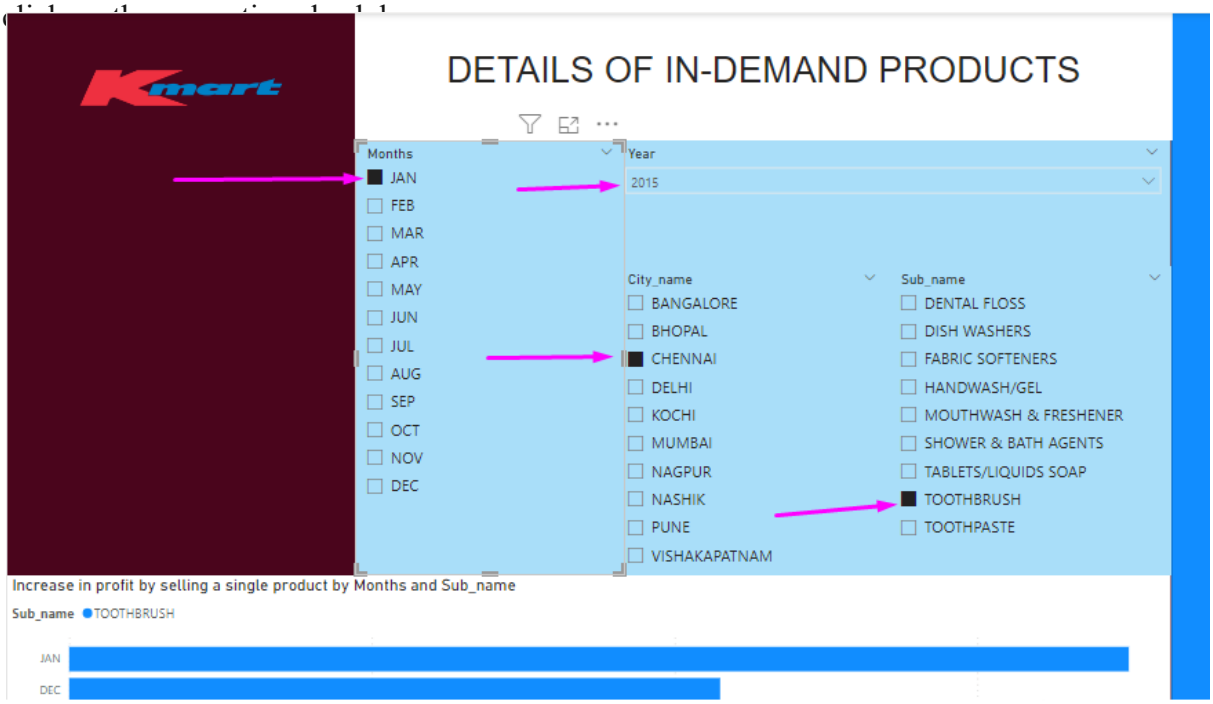
- Not in-demand products: Contains the detail about the product with maximum sales in a particular product type in a particular month and in a particular year. It also includes the supplier name of the particular product.

1.4.2 Navigating through the dashboard

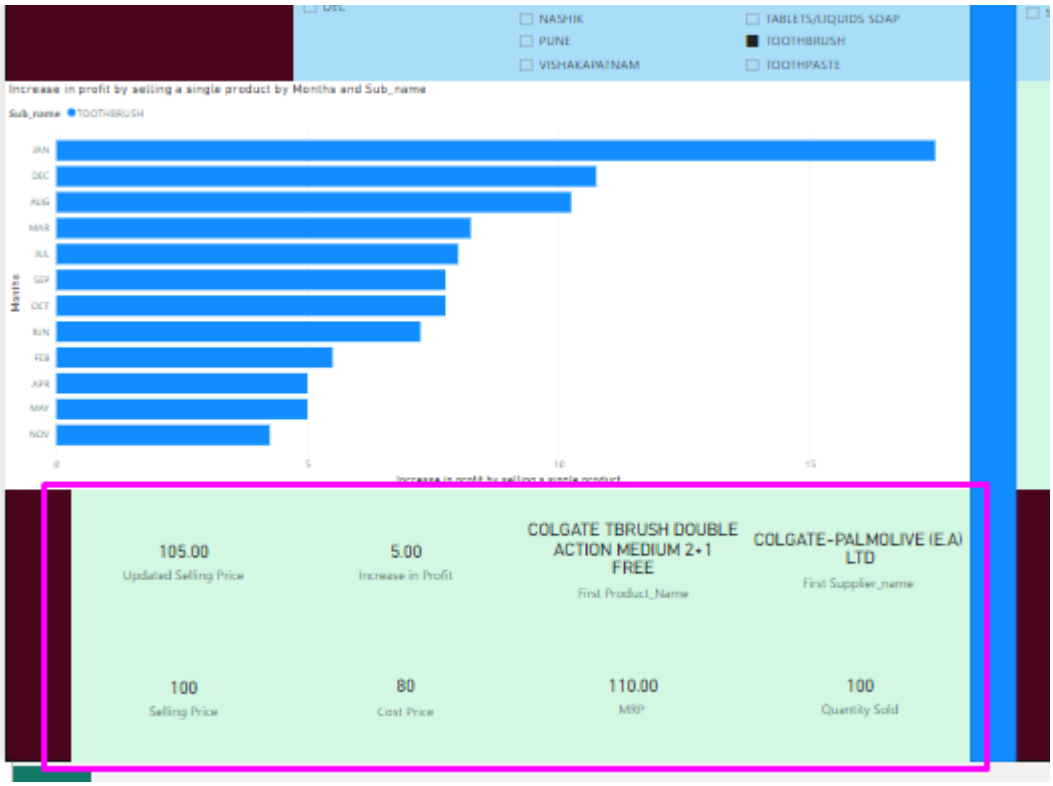
1) There are two sections in dashboard as shown in image and both these sections are independent.



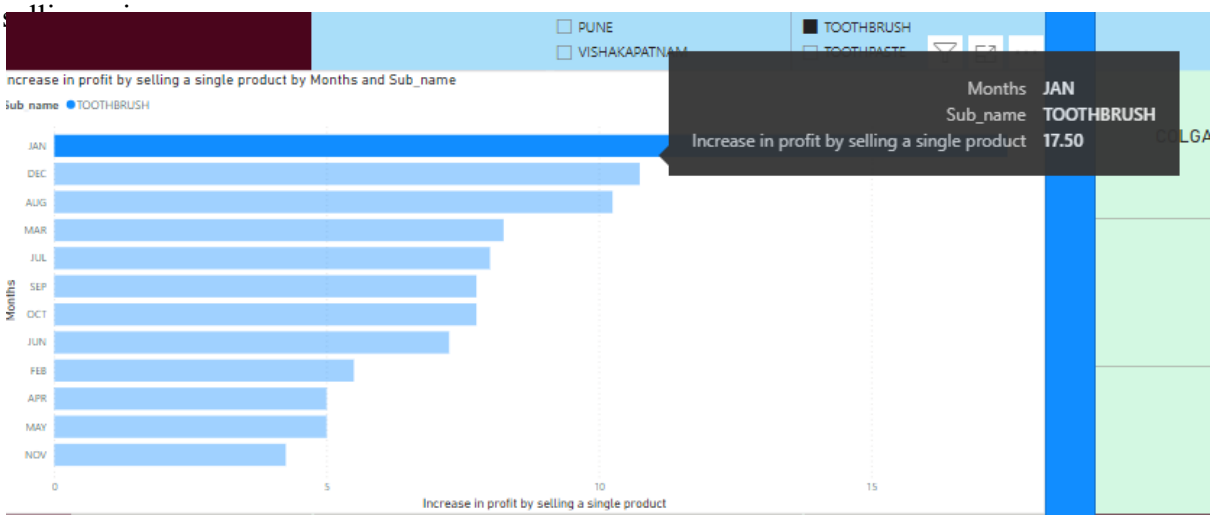
2) Say in Section 1 (In-Demand Products) if one has to find the product with maximum sales in Chennai at January 2015 and the type of products is tooth brush and then just



3) Now the results get updated automatically.



4) The stacked bar chart shows the increase in profit by selling a single product at new



5) One can follow the same steps for the section 2 (Not in-demand products) to get the required results.

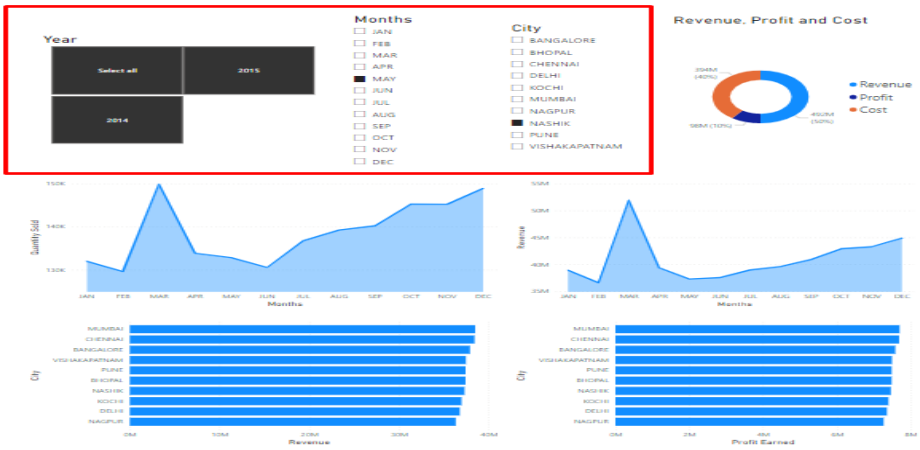
1.5 The Director of Operations Dashboard

1.5.1 About the dashboard

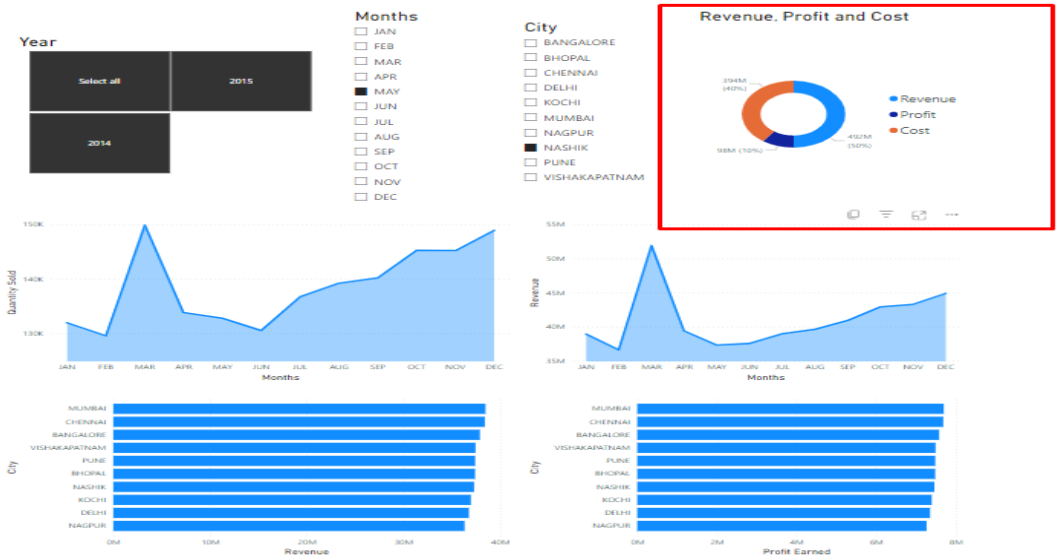
This dashboard helps the director of operations to analyze and compare the performance of each store in terms of quantity sold, revenue and profit generated.

1.5.2 Navigating through the dashboard

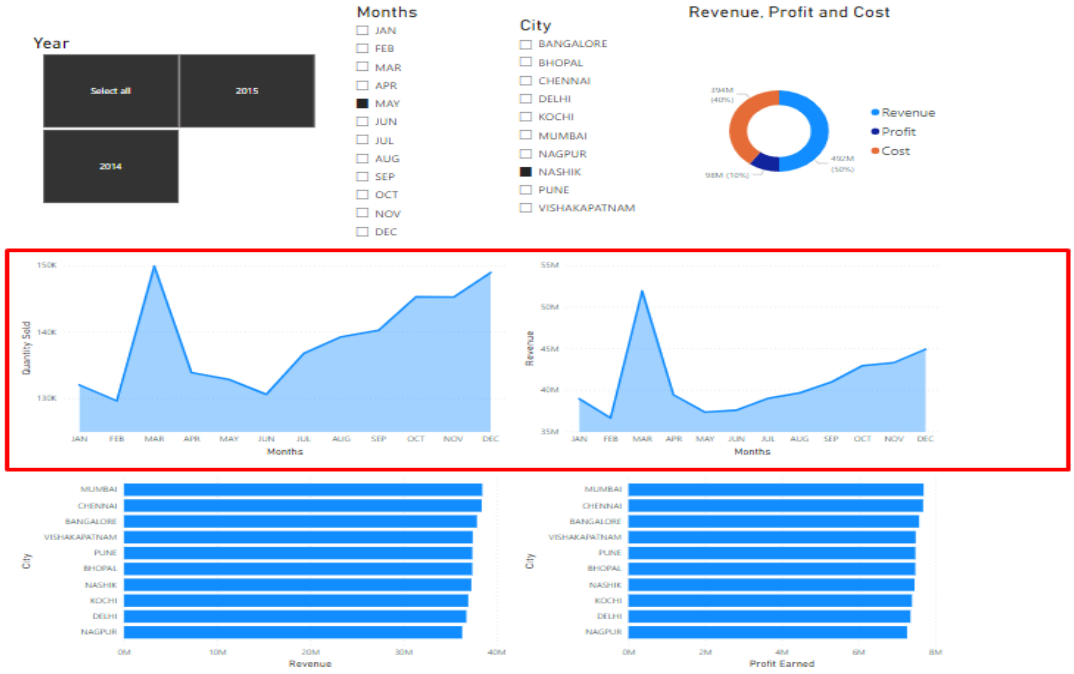
1) The three slicer highlighted helps one to select the year, month and city name to compare the performance of store.



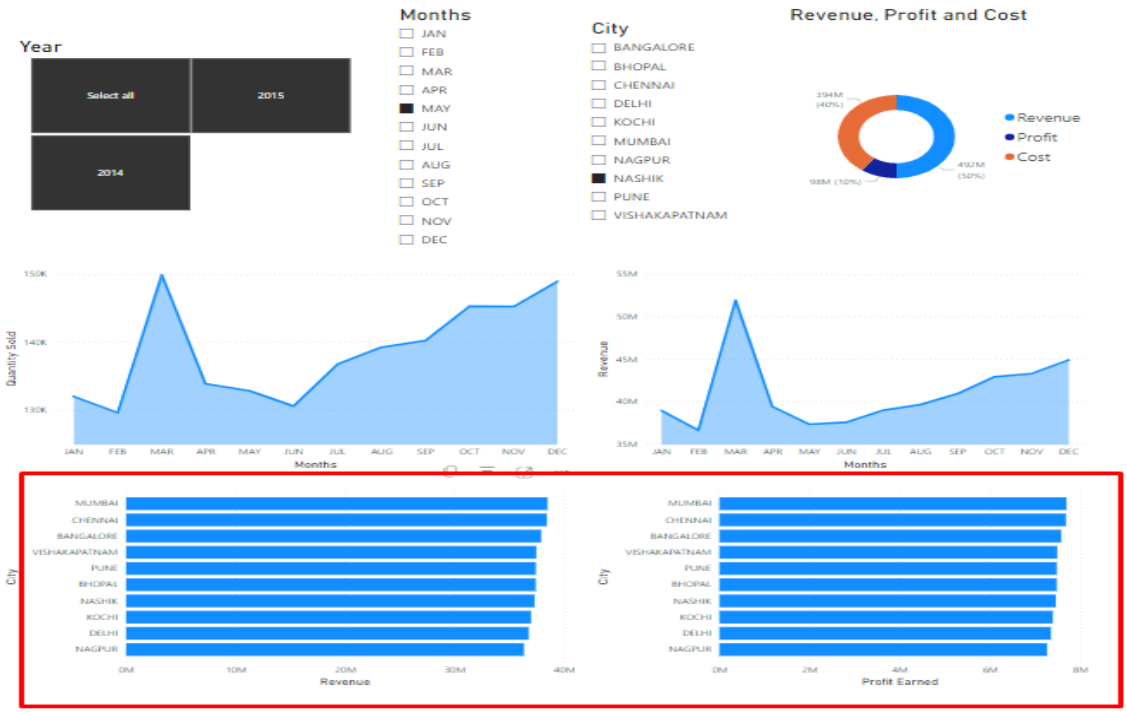
2) The donut chart on the right shows the cost of procurement, Revenue generated and the profit earned for the selected year, month and city.



3) The area chart on the left and right shows the quantity sold and revenue across



4) The stacked bar chart on the left and right shows the revenue and profit generated by stores for the selected year and month respectively.



2. Client end

2.1 Flow of project (sop)

a. Initiation – We acquainted ourselves with the problem statement and the requirements of our client
b. Planning – We divided the problem statement into multiple epics, further broke it into user stories and subtasks (Using AGILE TERMINOLOGIES). We did research and found the best softwares(platforms) to use for the problem statement, that satisfies the budget of our client. We made a workflow about how we are going to implement the project.
c. Executing – We linked AWS Services to Databricks for seamless integration and our analysts extracted insightful information from the data provided
d. Closing – Our product comprises of the graphical representation of the insightful data that was skilfully queried by our analysts.

2.1.1 Resources

a. Code documentation: - Our code is implemented using pyspark. It begins with connecting s3 buckets to data bricks using access keys generated from IAM. The next thing data ingestion from s3 to data bricks was done with help of read command. And further data analysing was done using data frame transformations
Click on Github for reference: Github
https://github.com/Aniishak/AAwDS3

b. Data documentation: Based on data provided, operations like data cleaning and normalization were performed on it.

c. Platforms and Tools: We did research and found the best software and tools that were best fit for delivering client requirements that satisfies the client budget.
    i. Storage: - AWS S3

ii. Processing and analysis: - Databricks
iii. Data visualization: - Power Bi

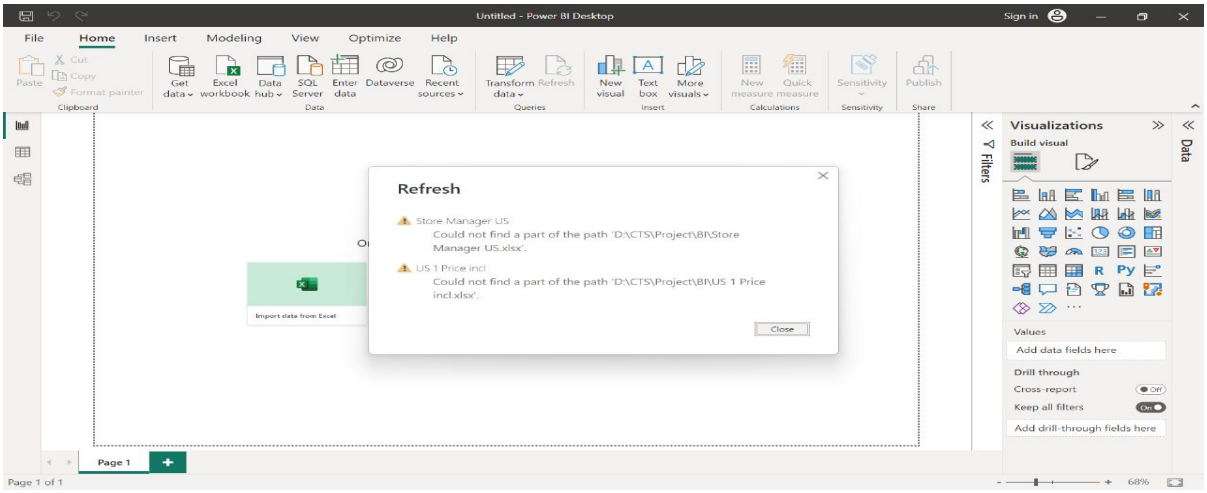
d. Work force: - Our workforce includes
    i. Scrum Master
    ii. Product Owner
    iii. Data Engineers
    iv. Data Analysts
    v. Power BI Developer

e. Key members (contacts)
    Swati thakre – 83xxxxxxxx
    A. Ajay sivaraman – 638xxxxx

3. SUPPORT TEAM

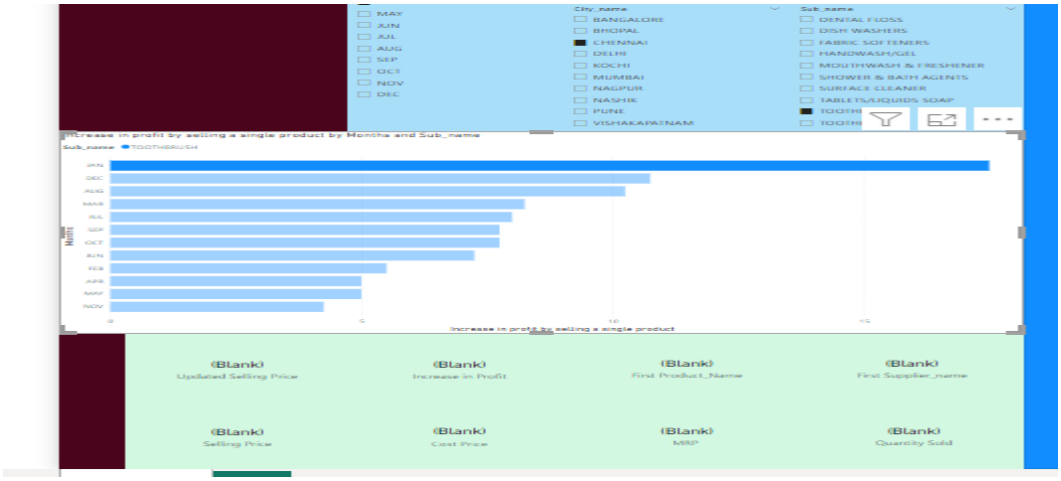
3.1 Troubleshooting

Problem 1: 'could not find a part of the path' appears while importing the dashboard



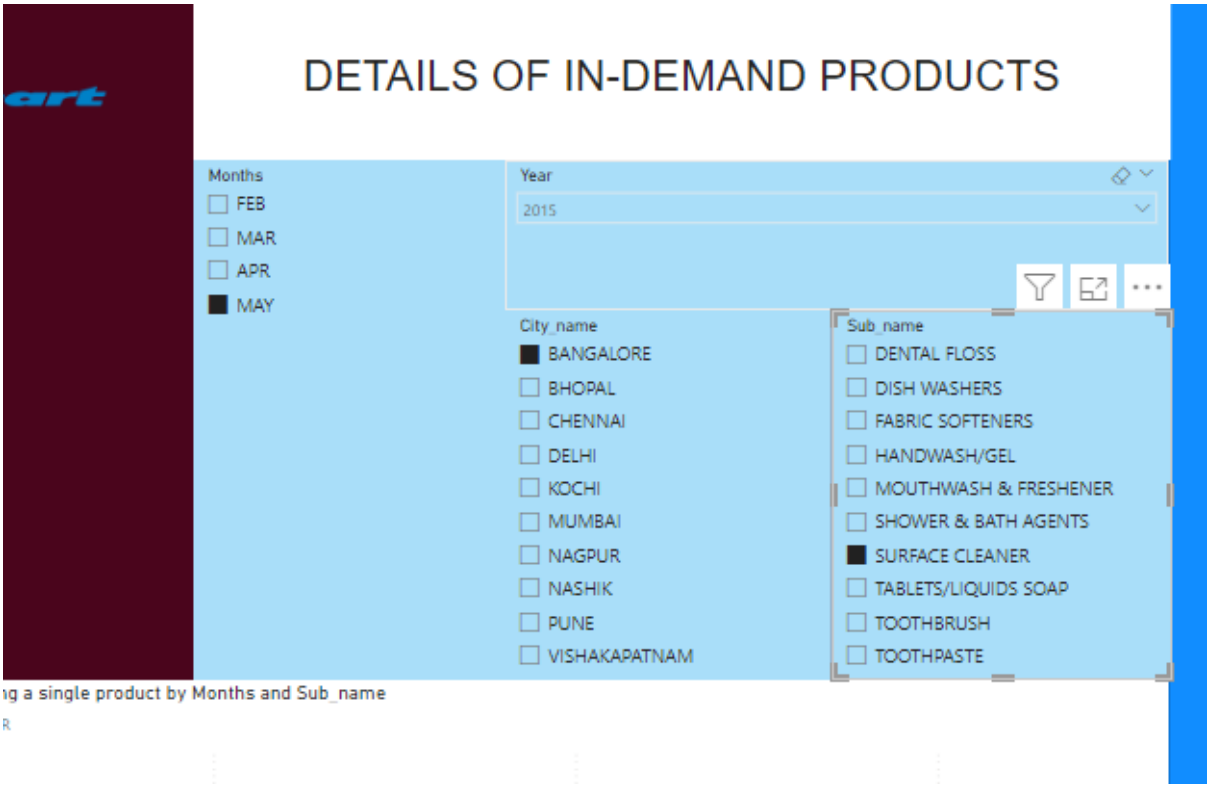
Solution: Just close the dialog box that appears on the screen everything will work fine (or) try to import the dashboard again.

Problem 2: All the results like product name, supplier name etc. displays blank instead of respective values.



Solution:You might have clicked on any of the bars in the graph so click at any random place inside the dashboard, everything will work fine.

Problem 3: For product type Surface cleaner, It allows to choose only 4 months to check the result.



Solution: This is not due to any technical problem; it just means that there are no sales of any of the surface cleaner in remaining 8 months of the year.

4. Security and Recommendations

4.1 Security Risks:

- Unsecured passwords
- Insufficient data encryption
- Vulnerable web applications
- Unsecured data storage

4.2 Recommendations:

To control these risks, the following recommendations were developed:

1. Strong password policy: Implement a strong password policy that requires users to use complex passwords and change them regularly.
2. Data encryption: Implement data encryption for sensitive data such as user data and payment information.
3. Web application security: Conduct regular vulnerability scans and penetration testing to identify and fix vulnerabilities in web applications.
4. Secure data storage: Implement secure data storage solutions such as encrypted databases or cloud storage.

4.3 Outcomes:

The recommendations were implemented successfully, and the following outcomes were achieved:

- All passwords were updated to meet the new password policy requirements.
- Sensitive data was encrypted using industry-standard encryption algorithms.
- Vulnerability scans and penetration testing identified and fixed several vulnerabilities in web applications.
- Secure data storage solutions were implemented to protect data.

4.4 Recommendations for Future:

To maintain the security of the system, the following recommendations are made:

1. Regular security audits: Conduct regular security audits to identify new risks and vulnerabilities.
2. On-going security training: Provide ongoing security training for employees to ensure that they are aware of security risks and how to avoid them.
3. Keep software up-to-date: Ensure that all software used in the system is up-to-date to prevent vulnerabilities caused by outdated software.

5. Software required

5.1 Software Name: Amazon S3

Description: Amazon S3 is a cloud-based object storage service provided by Amazon Web Services. It is designed to store and retrieve large amounts of data and objects, such as images, videos, and files, from anywhere in the world.

Role in Project: Amazon S3 was used as a storage solution for the project. It was used to store and manage various types of project-related data and objects, including

5.2 Software Name: Data bricks

Description: Databricks is a unified analytics platform that provides a collaborative workspace for data engineers, data scientists, and machine learning practitioners. It is built on top of Apache Spark and supports multiple programming languages, including Python, Scala, and R.

Role in Project: Databricks was used as a data processing and analytics platform for the project. It was used to perform various tasks related to data cleaning, transformation, analysis, and modeling.

5.3 Software Name: Power BI

We have used this Tool to do graphical representation of the analyzed data.

5.4 Software Name: Lucid Chart

Lucid Chart was used to represent normalized data and its schema

6. Decommissioning of a product

Standard Operating Procedure (SOP) for decommissioning a software product:

6.1. Notify stakeholders:

All stakeholders, including users and management, will be informed of the decision to decommission the software product. This will be done well in advance of the actual decommissioning to allow users to plan accordingly.

A mail from our end requesting the approval for decommissioning would be sent to each of these to you .

Point of contact would be appointed and available to you for all further communication regarding the decommissioning with signed document. These document's would then be validated by our legal team and by a hired lawyer agreed upon by both parties.

6.2 Upon validation further steps will take effect:

6.2.1 Determining the reason for decommissioning:

Before beginning the decommissioning process, it's important to establish the reason why the software product is being decommissioned. This could be due to the product becoming outdated, no longer being used by the organization, or being replaced by a newer product.

So the representative must have a document mentioning the reason for decommissioning ready.

6.2.2 Planning of the decommissioning process:

A detailed plan for the decommissioning process would be created, outlining the steps that need to be taken, the timeline, and the resources required. This plan would also include contingency measures in case of any unforeseen issues.

This plan must be approved by your representative.

6.2.3 Backup and archive data:

Before decommissioning the software product, all data associated with it should be backed up and archived. This data may be needed in the future for auditing purposes, legal reasons, or to provide continuity for any ongoing processes.

This backed up data would be given to the representative. The representative's signature would be required acknowledging the receiving of backed data.

6.2.4 Uninstall and remove the software:

The software would be uninstalled from all systems and removed from any servers or databases where it is installed. All associated files and data would also be removed.

6.2.5 Notify stakeholders of completion:

Once the decommissioning process is complete, all stakeholders should be notified of the successful decommissioning and any next steps that may be required.

A mail from our end would be sent to the stakeholder's informing them of the completion.

6.2.6 Conduct post-decommissioning review:

After the decommissioning is complete, a post-decommissioning review should be conducted to evaluate the effectiveness of the process and identify any areas for improvement.

6.3 Security aspects to ensure while decommissioning a software product.

6.3.1 Data backup and retention: Before decommissioning the software, it is essential to back up and retain all data associated with the application. This ensures that critical information is not lost and can be accessed if needed for legal or regulatory purposes.This would be done by us and the back-up would be given to the representative.

6.3.2 Data destruction: All data and information associated with the software product would be securely destroyed or deleted after the software has been decommissioned. This includes any backups, copies, or versions of the data that may exist.

6.3.3 Access control: Access to the decommissioned software product would be restricted to authorized personnel only. This ensures that sensitive data or information is not accidentally or intentionally accessed, modified, or deleted. As to who these authorized personnel would be, should be decided by the stake holder's at the time of decommissioning.

6.3.4 Configuration management: All configurations of the decommissioned software product would be removed, and any security-related configurations or settings would be reviewed to ensure that they are no longer active and do not pose any risk to your organization.

6.3.5 System hardening: The systems and servers on which the software product was installed would be hardened to prevent any unauthorized access or attacks. This may include applying security patches, disabling unnecessary services or ports, and changing default passwords.

6.3.6 Audit trails: An audit trail of all actions taken during the decommissioning process would be maintained to provide a record of who accessed the system and what actions were taken.

6.3.7 Communication: Stakeholders: including users and management, would be informed of the decommissioning process and the security measures being taken to ensure that sensitive data is protected. This process would be done through mail.

By considering these security-based aspects, organizations would ensure that decommissioning the software product is carried out in a secure and controlled manner.

6.4 Responsibilities of our support team during decommissioning of a software product:

6.4.1 Communicate with stakeholders: The support team would communicate with all stakeholders, including users, management, and other relevant teams, to inform them of the decommissioning process and answer any questions or concerns they may have.

6.4.2 Plan the decommissioning process: The support team would work with other teams to plan the decommissioning process, including identifying the systems and applications that need to be decommissioned, determining the timeline, and identifying any potential risks or challenges.

6.4.3 Provide technical support: The support team should provide technical support to other teams during the decommissioning process, including assisting with the backup and retention of data, removal of the software from systems, and ensuring that all configurations and settings are properly removed.

6.4.4 Document the process: The support team should maintain accurate documentation of the decommissioning process, including any steps taken, decisions made, and any issues or challenges encountered. This documentation can be used to evaluate the process after it is complete and for future reference.

6.4.5 Test the decommissioning: The support team should work with other teams to test the decommissioning process to ensure that it is effective and that all data and systems are properly secured.

6.4.6 Provide post-decommissioning support: The support team should continue to provide support to other teams after the decommissioning process is complete, including answering questions or addressing any issues that may arise.

By taking on these responsibilities, the support team can ensure that the decommissioning process is carried out smoothly, with minimal disruption to the organization, and that all sensitive data and systems are properly secured.