E-Commerce Customers Behavior Research Using Power BI - Cohort Analysis: A Case Study

Ashok Reddy Thontla, Research Scholar, K L Business School, KLEF (Deemed to beUniversity), Email: ashokthontla@gmail.com

Dr.N.Bindu Madhavi, Associate Professor, K L Business School, KLEF(Deemed to be university), Email: dr.bindumadhavi@kluniversity.in

Abstract

In the current world the data is growing rapidly on day by day, there must be an essential to analyze the huge volume of data using business intelligence tools and techniques. BI tools and techniques have the capability to turn this huge raw data into a business information to fulfill all the business needs and their problem resolution. BI is the wide area to find the solutions for business problems it may be Sales, Insurance, Finance, HRM, SCM, Procurement, etc.

Most of the IT people are aware of the BI tools and their technologies there are many of the reporting tools are available in the market before going to the tools, what exactly the reporting does mean, the reporting it contains a collection of pages, and each page contains the information, and it would be captured and presented to the higher people for the decision-making purpose.

Introduction

Many reporting tools which can be used to designed/developed of the reports the given list I have presented having various reporting tools are available in the market in which we can categorize few of them are reporting tools and few of them are data visualization tools.

if you see in the list d3js is a data driven dashboard for JavaScript which is the data visualization tool in the similar way we have Power BI, Spot fire, SAP Lumira and Tableau so the Power BI and Tableau are in leading position. Also we have traditional reporting tools as such as SAP BO, SSRS, SSAS ,Micro Strategy , Pentaho which can present the data in the form of paginated reports and these doesn't have much data visualization capabilities to present the data into the viewers so that's where visualization come to the picture and moreover these are traditional reporting tools allows to bring the data only for a few of the data sources not from all the data sources.

We have a different data sources could connect with on premises data sources in which there are some files which are existed in your system it may be Excel, CSV , and XML file present in your system and are trying to pull the data from those heterogeneous data sources to be populating data to create a report.

These data visualizations have all the capabilities where it allows various heterogeneous data sources in which you could connect not only for a particular data source system it allows all the five data sources are available.

Research Paper   UGC CARE Listed (Group -I) Journal

> ➢ On Premises Data Source
> ➢ Online Data source
> ➢ On Premises Database
> ➢ Online Database
> ➢ Online Data services

On premises data sources is the source of data that resides in Excel, CSV, and XML file's where you are trying to connect from the offline physical system.

Online data sources are the source of data that resides in Google cloud or OneDrive system where you are trying to connect from the online

On premises database is the source of data that you are trying to connect with SQL Server Oracle, Postgrel, SAP Hana, SAP BW so all these are the heterogeneous databases where you could connect from the offline database

Online database is the source of data that you are trying to connect with Microsoft azure, Azure HD insight and AWS Redshift through the online databases.

Online data services are the source of data that you are trying to connect with online data services in which Facebook analytics, Google analytics and Adobe analytics.

Power bi is a collection of software and services where you could connect with any kind of source system to bring the data to do the data analysis and data modeling and apply all sort of business transformations to create the interactive report in which it can be presented to the higher people for effective decision making.

Power BI is data visualization tool in which it is a graphical representation of the data where you get the data insights for Powerful decision making data visualization is a part of business intelligence to convert the raw information into a business or meaningful information. We have many traditional reporting tools in the market which can design / develop the reports, interact the reports to provide the business solutions such as:

Sales, Insurance, HR, Marketing, Life science, Education, Procurement and Supply Chain management problems. Business Intelligence can provide the solutions for business problems to understand your business data in terms of for data discovery, data analysis, data visualization and data exploration.

Review of Literature:

1. Alam et al., 2007a

   S.S. Alam, A. Khatibi, H. Ismail, B. Ismail, Hishamuddin, S. Ahmad

   Factors Affecting E-Commerce Adoption in the Electronic Manufacturing Companies in Malaysia

International Journal of Commerce and Management, 17 (1/2) (2007), pp. 125-139

View in ScopusGoogle Scholar

2. Almoawi, 2011

   Almoawi, Abdul Rahman Nasser A. 2011. E-Commerce Adoption Among Small and Medium Enterprises in Saudi Arabia. PhD thesis, Universiti Utara Malaysia.
   Google Scholar

3. Alzougool, 2008

   Alzougool, B., Kurnia, S. 2008. Electronic Commerce Technology Adoption by SMEs: A Conceptual study. 19th Australian Conference on Information System 3-5 Dec 2008, Christchurch Australia.
   Google Scholar

4. Cavana et al., 2011

   Cavana, R.Y., Delahaye, B.L. ,Sekaran, U. 2011. *'Applied Business Quantitative and Qualitative Methods*,' Sidney: John Wiley & Sons.
   Google Scholar

5. Chang., 2006

   Chang. S.L. (2006). Organizational, technological and environmental determinants of E-commerce adoption in SMEs in Taiwan. Unpublished PhD thesis, Lyn University, US.
   Google Scholar

6. Choi et al., 1997

   Choi, S., Stahl, D., Whinston, A. 1997. *The Economics of Electronic Commerce*. Indianapolis : Macmillan Technical Publishing.
   Google Scholar

7. Davis, 1997

   W.R. Davis

   The FC1 rule of identification for confirmatory factor analysis: A general sufficient condition

   Sociological Methods & Research, 21 (4) (1997), pp. 403-437

   View article

   CrossRefGoogle Scholar

8. Elmazi et al., 2011

I. Elmazi, H. Vukaj, E. Gega, L. Elmazi

Information technology and its effects in SME

The case of Albania. International Journal of Management Cases, 13 (4) (2011), pp. 291-298

View article

CrossRefView in ScopusGoogle Scholar

9. Gilaninia et al., 2011

S. Gilaninia, S.Y. Danesh, M. Amiri, S.J. Mousavian, B. Eskandarpour

Effective Factors on Adoption of E-Commerce in SME Cooperative

Interdisciplinary Journal of Contemporary Research inBusiness, 3 (6) (2011), pp. 144-161

Google Scholar

10. Grandona, 2004

E.E. Grandona, J.M. Pearson

"Electronic commerce adoption: an empirical study of small and medium US businesses", Information & Management, 42 (1) (2004), pp. 197-216

11. Harrington, 1996

L. Harrington, G. Reed

Electronic Commerce Finally Comes of Age

The Mckinsey Quarterly, 2 (1996), pp. 68-77

Google Scholar

12. Kearney, 2013

Kearney, A.T. 2013. Global retailers: cautiously aggressive or aggressively cautious? https://www.atkearney.com/consumer-products-retail/global-retail-development-index/full-report/-/asset_publisher/oPFrGkbIkz0Q/content/2013-global-retail-development-index/10192. Accessed 22 Nov 2013.
Google Scholar

13. Lertwongsatien, 2003

C. Lertwongsatien, N. Wongpinunwatana

E-commerce adoption in Thailand: An empirical study of small and medium enterprises (SMEs)

Journal of Global Information Technology Management, 6 (3) (2003), pp. 67-83

View in ScopusGoogle Scholar

14. Lip-Sam, 2011

T. Lip-Sam, L. Hock-Eam

Estimating the determinants of B2B e-commerce adoption among small & medium enterprises

International Journal of Business and Society, 12 (1) (2011), pp. 15-30

View in ScopusGoogle Scholar

15. MacGregor and Vrazalic, 2008

R. MacGregor, L. Vrazalic

A profile of Australian regional SME non-adopters of ecommerce

Small Enterprise Research, 16 (1) (2008), pp. 27-46

View article

CrossRefView in ScopusGoogle Scholar

16. Pfeffer, 1977

J. Pfeffer, H. Leblebici

Information technology and organizational structure

Pacific Sociological Review, 20 (2) (1977), pp. 241-261

View article

CrossRefView in ScopusGoogle Scholar

17. Premkumar, 1995

G. Premkumar, K. Ramamurthy

The role of organizational and inter-organizational factors on decision mode for adoption of inter-organizational systems

Decision Sciences, 26 (3) (1995), pp. 303-336

View article

CrossRefView in ScopusGoogle Scholar
18. Reich, 1990

B.H. Reich, I. Benbasat

An empirical investigation of factors influencing the success of customer-oriented strategic systems

Information Systems Research, 1 (3) (1990), pp. 325-347

View article

CrossRefView in ScopusGoogle Scholar
19. Rogers, 1962

Rogers, E.M. 1962. *Diffusion of innovations* (1st ed.). New York: Free Press.
Google Scholar

20. Saffu et al., 2012

K. Saffu, J.H. Walker, M. Mazurek

Perceived Strategic Value and e-Commerce Adoption among SMEs in Slovakia

Journal of Internet Commerce, 11 (1) (2012), pp. 1-23

View article

CrossRefView in ScopusGoogle Scholar
21. Sekaran, 2009

Sekaran, U. 2009. *Research Method for Business, 5th ed*, *UK:* John Wiley & Sons, Ltd.
Google Scholar

22. Seyal, 2003

A.H. Seyal, M.N.A. Rahman

A preliminary investigation of e-commerce adoption in small & medium enterprises in Brunei

Journal ofGlobal Information Technology Management, 6 (2) (2003), pp. 6-26

View in ScopusGoogle Scholar
23. Shah Alam et al., 2011

S. Shah Alam, M.Y. Ali, M.F. Mohd Jani

Research Paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal

An empirical study of factors affecting electronic commerce adoption among SMEs in Malaysia

Journal of Business Economics and Management, 12 (2) (2011), pp. 375-399

Google Scholar

24. Soh et al., 1997

C. Soh, Q.Y. Mah, F.J. Gan, D. Chew, E. Reid

CrossRefView in ScopusGoogle Scholar

25. Zwass, 1996

V. Zwass

"Electronic commerce: structures and issues"

International Journal of Electronic Commerce, 1 (1) (1996), pp. 3-23

Google Scholar

Different visualization tools in the market

- ➢ Power BI
- ➢ Tableau
- ➢ Qlik View/Qlik Sense
- ➢ Spotfire
- ➢ SAP Lumira

All the above tools have unique capabilities and functionalities, but Power BI and Tableau are in leading position in the Data Visualization world.

Power BI and Excel both are powerful together to fulfill any business solution but excel is an adhoc tool whereas having data limitation. But both are having similar functionalities. Below are the various functionalities of the Power BI.

- ➢ Data Discovery
- ➢ Data Analysis
- ➢ Data Visualization
- ➢ Data Exploration

Data Discovery: Is to bring the data from various source systems such as on premises, online.

Data Analysis: Is to build the data model and analyses data based on various characteristics

Data Visualization: Is to create the reports and dashboards using various inbuilt and custom visuals.

Data Exploration: Is to explore the data using google Map's and filled maps'

There are different roles in the Power BI

- ➢ Power BI Designer
- ➢ Power BI Developer
- ➢ Power BI Report Consumer and
- ➢ Power BI Admin

1. Power BI Introduction

You will learn all about PBI desktop,PBI pro,PBI mobile and more because PBI is not a single tool it's a set of combined tools and in this course you will learn all to go from zero to creating your own PBI projects we will learn all this step-by-step in an interactive course project the project starts with PBI desktop which is a local application which you can download to your Windows computer,PBI desktop is all about data preparation of analysis and visualization and we will cover all these three parts in a lot of detail after PBI desktop we'll have a look at PBI Pro which allows us to bring our data from the local computer to the cloud using power Pro license.

And you can collaborate in your projects with your colleagues and you can share projects within and outside your organization will then continue with PBI mobile its name indicates a mobile application which allows you to access your information from anywhere from your mobile device after that we'll dive into more advanced features of PBI for example we will learn how to create own custom visuals PBI ships with a lot of beautiful visuals but if that's not enough also you can dive deeper and create visuals exactly according to your needs on your own.

Let's get started to dive right into the insights of PBI and to do that we should have a look at the most important question first what is PBI actually on the official website you can find this statement here telling you that power bi is a business analytic solution it allows you to visualize content and to share information and that it can be connected to hundreds of sources and you can use it to create dashboards and reports that's a nice statement indicating that PBI is a quite powerful but I guess it's a bit too complex at this stage because in simpler words power bi is a BI a business intelligence allowing you to analyze visualize and share that's it in the end and there are some thing is that PBI is a BI for everyone so you don't need any prior experience to work with PBI now.

We will dive into all its core features throughout the course but at this point we should understand how PBI is organized and structured to allow everyone to become a BI expert for that we should have a look at the three core areas of PBI the first core area is the preparation of data and analysis so anything related to tables you could say preparation means connecting PBI to different source files and to work on that raw information this can mean that you want to work on columns delete columns for example or you might want to change the format from a text to a number the goal of preparation is to have shaped or prepared which you can then use to perform your analysis analyzing information in PBI could mean simple tasks like calculating averages or more complex calculations we'll have a look at throughout the course with the data related part finished we can continue with the next core area visualizations because you can of course present

as simple tables this is also possible and PBI but you can also create lots of different visuals this can be line charts column charts bar charts or more advanced visuals like maps for example so that's the second core area we now saw that we can work and that we can present our results and this brings us to the third and last core area of PBI is collaboration and sharing because you can work on your own in PBI this is possible but you can also work together in teams and you can also share your results then within your organization or even outside your organization this is collaboration and sharing so these are the three core areas making PBI such a powerful BI the last question now is how does PBI do this though which do I have to achieve all these goals

Well this brings us to the core PBI it because PBI isn't a single it's a set of multiple combined groups interacting with each other and these core are PBI desktop,PBI service or Pro and PBI mobile let's get started with PBI desktop is a local application this means it's a program you can download for free to your computer with one single restriction though it's only available for Windows machines you have a very strong covering many of the core areas we just talked about and you can work on data sets this means the initial data preparation we just talked about so the connection to the source files empty work on that information and also the data analysis part meaning that you apply your own calculations to the information to get the insights and also create visuals in PBI desktop.

As I said we have line, bar, column charts Maps even combined so many different ways to visualize information and to bring more structure into your visuals PBI has the so called report feature a report simply allows you to organize your visuals on different pages so you have one page for example with revenue data a second page with cost data just to give you a simple example here and in the end all these different pages are summarized in one report this is what about in a nutshell as you can see a very powerful already it has one core restriction though it can be downloaded to your machine and therefore you are kind of alone when you work in for certain projects but typically if you work in an organization you are not alone and you're working in teams.

And this brings us to the second core PBIas PBI service or pro is a cloud application this means you just need a account and then you can use it to log in which means it's not limited to Windows only you only need a browser to well access this website what can we do in PBI service it is not the ordinary to work on datasets that's part of PBI therefore we don't have these strong data preparation and analysis capabilities which has talked about in and can be used to create visuals and reports though this is possible but it works basically the same way as it does in desktop therefore in this course we will focus on PBI desktop when it comes to visuals and reports and not dive too deep into this area in PBI service because there is actually another part where PBI service really shines and this is creating dashboards in our summary pages you could say one single page with the core visuals of the report this could be a page you present your executive to give them.

An overview of the most important numbers from the dashboard you can then dive into the report and into single visuals if you want you but the dashboard is the first thing you see now as I talk about presenting dashboards to your executive this brings us to the other really strong part of service, this is collaboration and results sharing because service is the need when it comes to any kind of interaction with other people be it inside your team beat inside your organization or even outside your organization you can use service to collaborate on your object and to share the final

results with other people and this sharing part also brings us to the last core mobile as the name indicates Mobile is a mobile app which you can download from the Google Play Store or the App Store if you have an Apple device and it simply allows us to access our information our dashboards our reports from anywhere and this is not core bi that was the fury part and throughout this course we'll dive into all these areas we'll see how these topics interact in a real interactive project.

2. Parts Of the Power BI
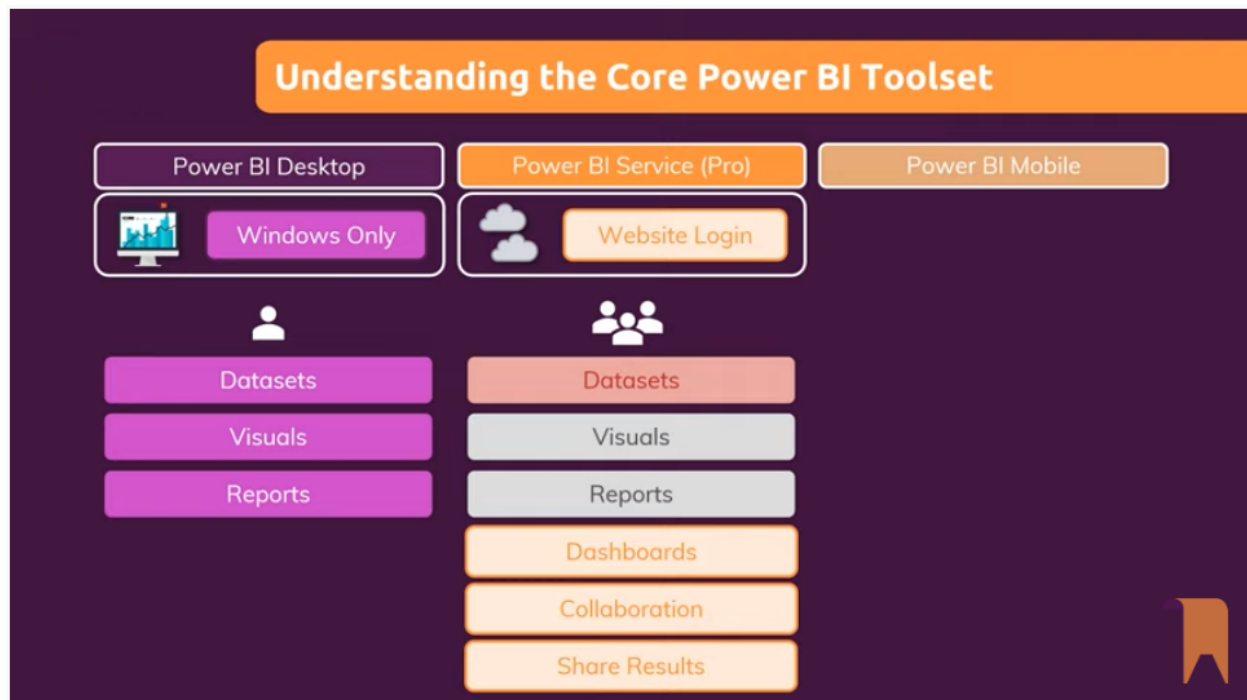
- ➢ PBI Desktop
- ➢ PBI Service
- ➢ PBI Mobile

PBI Desktop: Is a windows desktop application which allows the users to connect with any kind of source system may be on premises data sources, online data sources, on premises databases, online databases, and online data services to import the data into the PBI, there after build the data model where you could connect with multiple tables in an organized manner by following the relationship models. Apply all sort of transformations to convert the raw data into business information. Design or develop the reports by consuming the data from PBI dataset. Creation of the reports can be done based on multiple visuals of representing the data then finally publish the report to the PBI service.

PBI Service: Is a collection of software and services in which the reports can be published to the service and shared among multiple users with in or outside organization. Reports can be

refreshed on day to day using the gateway configuration. PBI service should have PRO or Premium account to publish, and collaboration of the reports and reports can be scheduled using scheduled options in the gateway configuration. Subscriptions and alerts can be created with in the PBI service. Dashboard can be created on top of the reports which contains collection of information. All the security part can be managed in the PBI service.

PBI Mobile: It allows the users to connect to PBI service and to interact with the reports' mobile is available for a different version of devices windows, android, iPhone, IPOD. Data can be previewed and interacted anywhere else in the globe using mobile devices.

it allows you to work on data to create reports and publish the reports and a lot more this sounds really great, but we have to take a closer look to understand how this works, so we already know that PBI is not a single tool but consists of multiple tools additionally we have that basic building blocks of the PBI.

3. Building blocks of the Power BI

Power BI Building Blocks – 5 Major Components of Power BI
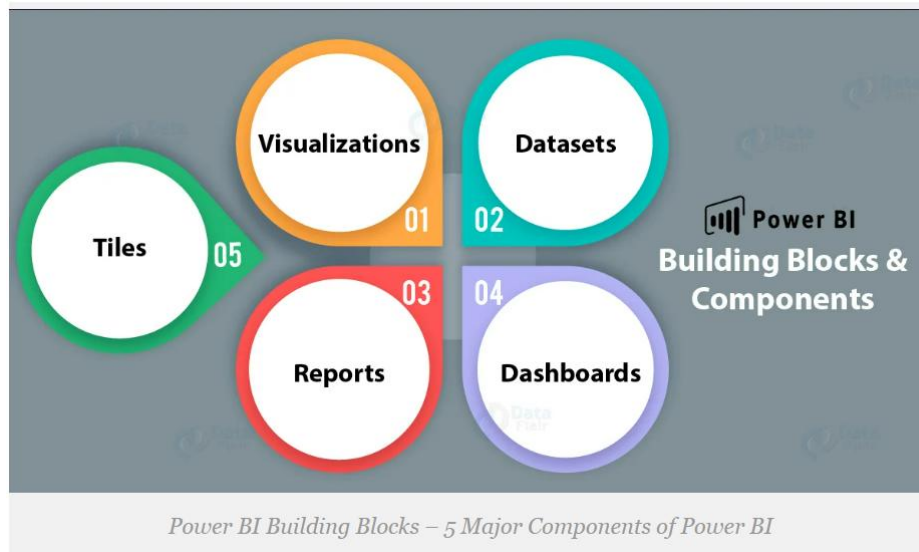
➢ Visualizations
➢ Dataset
➢ Report
➢ Dashboard
➢ Tiles

Dataset : Describes the collection of data queries which gets the data from homogeneous or heterogeneous sources to import the data into model and apply all sort of business transformations to convert raw data into meaningful information. Dataset which consists of data model where we can organize the data using data warehousing models ex: star schema or snowflake schema.

There is different sort of transformations and DAX calculations which can apply on the data.

➢ Data Conversions
➢ Data Sorting
➢ Data Scrubbing

Research Paper

- ➢ Data Cleansing
- ➢ Data Merging
- ➢ Data Categorization
- ➢ Data Validation

Report: Is a collection of visuals and each visual contains collection of information which can be present to the users for effective decision making. Reports can be shared to the multiple users for data analysis.

Visualizations: Is a collection of visuals and each visual is a graphical representation of the data in which embedded the visuals into the report.

Tiles: Each individual visual called as a single Tile, which contains graphical representation of the data. Collection of tiles would be blinded to the dashboard.

Dashboard: Is a collection tiles which can be organized in a single page, whereas report contains collection of pages to present the data from various domains. Here each domain data can be presented in one single dashboard.

4. Different roles in the Power BI

They are different roles in the PBI so let's say once if you are hired for the PBI position what role you are looking for, is it for PBI designer, PBI developer, PBI report consumer and PBI admin.

what are the functionalities and responsibilities of all these various system let's say PBI designer typically what would do, pull the data I mean would connect to the various Source systems and pull the data to perform different sort of transformations whereas data sorting , data conversion , data scrubbing , data combining, data merging and data validation so all these are the different sort of data transformations which you could apply on your data to transform or to translate your raw data into a meaningful information and finally you could bring the data into your power bi so where you could create or design the reports which is ready to present to the higher officials.

 The PBI report consumer is a final user who would not be part of development and they would be part of requirement gathering they are the final users who will consuming the report they are higher officials who will be part of the business and they would be able to plan out of your data.

PBI report developer create a report and embed this report to any other applications. Which means these reports can be embedded to the applications if you are the power bi developer you would be knowing all the coding part whereas knowing a Python language or R language

PBI admin is to perform the actions to provide the access to the reports and will manage all set of reports in the server and if any kind of access related or if any kind of user authentication related admin would come to the picture.

Various Power BI Architecture's

PBI has various architectures depends on usage and volume of data and users it varies from one to other.

- ➢ Personal BI
- ➢ Small Team Collaboration
- ➢ Large team collaboration and distribution

- ➢ Enterprise Level Collaboration and Distribution
- ➢ Centralized Corporate Reporting and Data Warehousing
- ➢ Prototyping Activities and Sharing
- ➢ Organizational Embedded Analytics
- ➢ External Embedded Analytics
- ➢ On-Premises (Non-Cloud) Reporting Portal
- ➢ Workflow Automation Updating Data from Within Power BI

1. Personal BI

The below architecture comprises on Personal BI capabilities, which consists of various parts

PBI desktop allows to bring the data from various on-premises and online data sources.

Pulling the data from various queries and data mashups apply the data transformation to turn the data into meaningful information. Creation of the reports on the loaded data into the PBI system there are after publish the reports to PBI services and configure the PBI gateway to refresh the latest data from the on premises source system to PBI service. Power BI service comprises the various building blocks of PBI along with creation of the My workspace where the reports can be published to the private space.

PBI Mobile is one of the major part where the user could log on to their authenticated credentials to interact with the various reports and dashboards to perform key decisions to improve the organization growth.
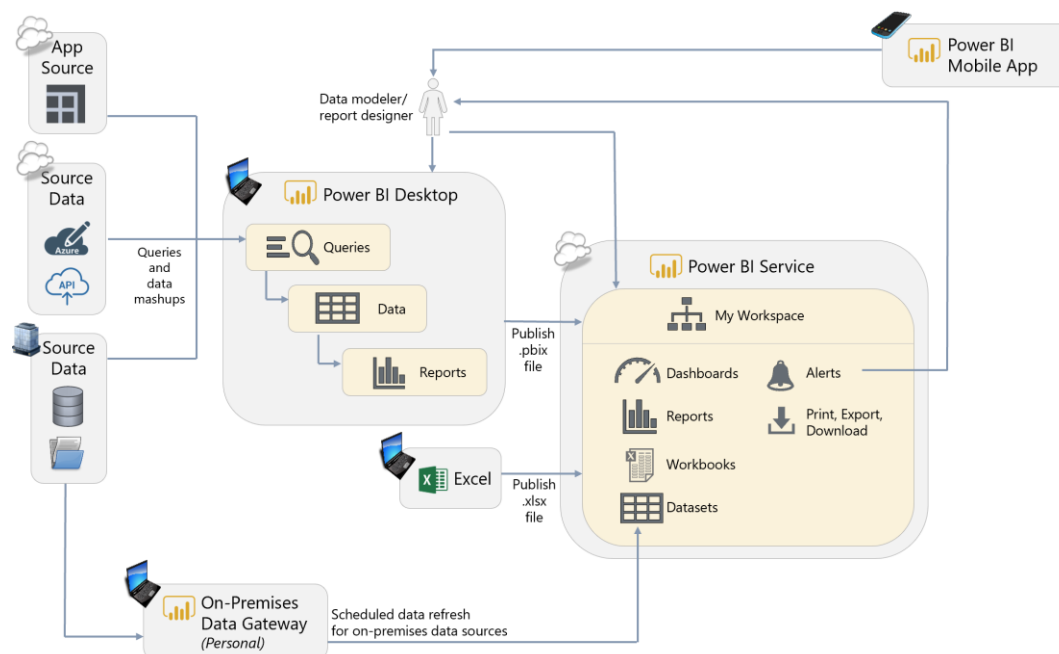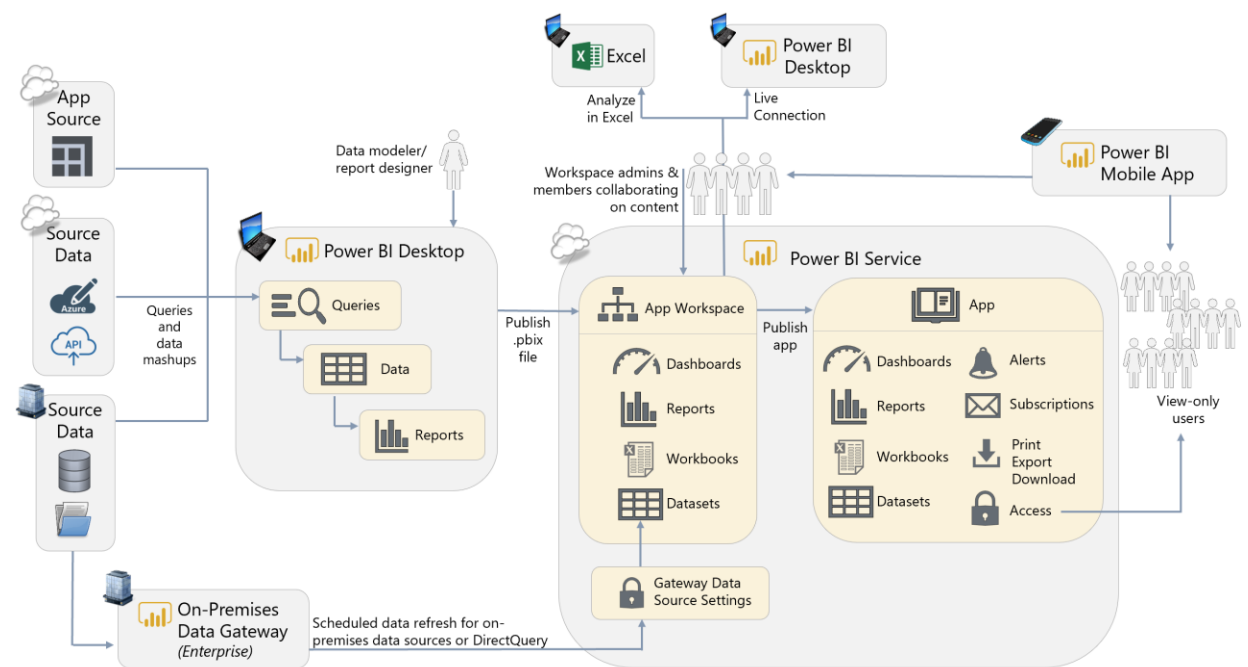

2. Small Team Collaboration

The below architecture comprises same level of components in addition to that App workspace would be there to publish the reports, the team can collaborate and work together on the report's interaction. Alters, subscriptions and dashboards would be created in the PBI service.
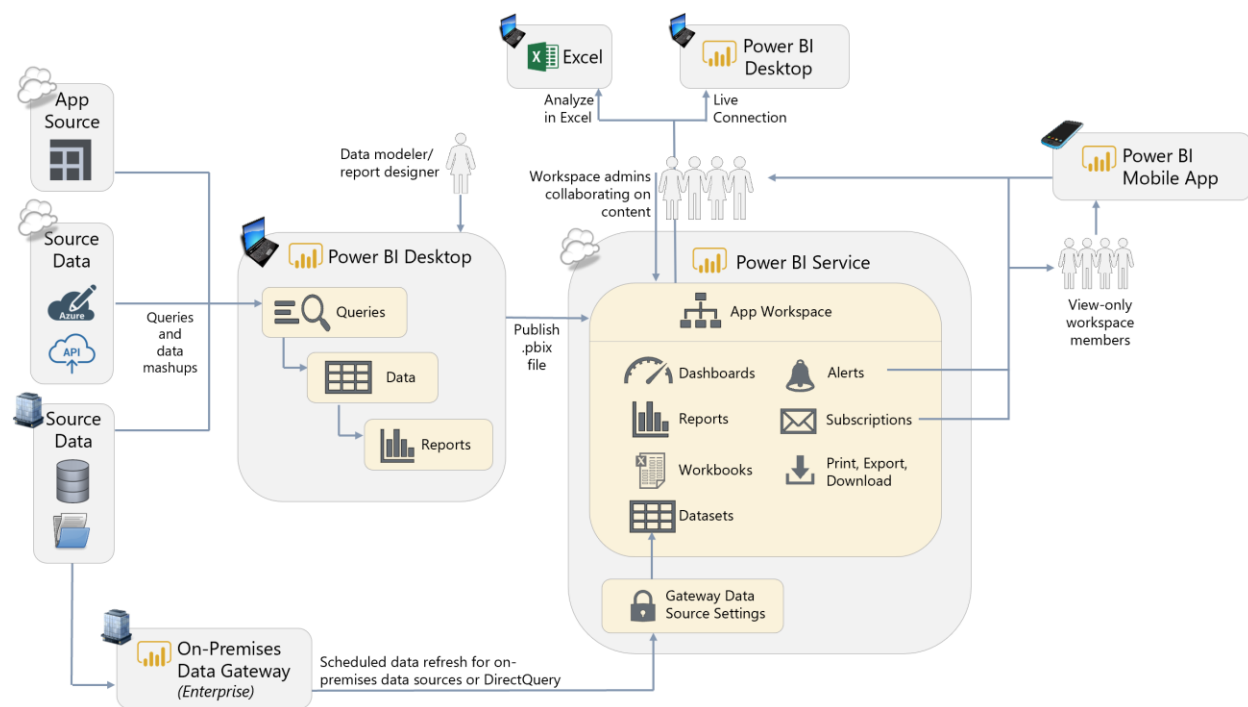
3. Large team collaboration and distribution

PBI architecture consists of different components which can distribute the reports to the huge volume of users with large data.

PBI Components:

- ➢ Data Sources
- ➢ PBI Desktop
- ➢ PBI Service
- ➢ On-Premises Data Gateway
- ➢ Power BI Mobile
- ➢ Analyze in Excel

PBI gets the data from various source systems may be on premises, online data sources and databases. Which perform the data mashups, prepare the data queries to analyze and build the data models for creation the reports there after need to publish them to the PBI service. Before going to publish the reports make sure it can be published to My workspace or App workspace.

Whatever the reports published under my workspace can be visible to you only and others can't see your reports until you share them whereas the reports under app workspace can be shared and collaborated among the team major distribution of the reports can be Published through app workspace.

Data need to be refreshed when ever there is a data change in the backend data source, to refresh the data gateway need to be configured. Make sure Gateway need to be installed in the dedicated server and the services should be up and running to refresh the data from the dataset.PBI Gateway can act as bridge between on premises data sources and databases to the cloud (SaaS), which can transfer the secure data through azure bus. PBI reports can be shared to the users through PBI App. App is a collection reports which can be embedded into one group, whoever are entitled to access the reports.

Power BI architecture can vary based on organization need and set of the people who are going to access them.

4. Enterprise Level Collaboration and Distribution
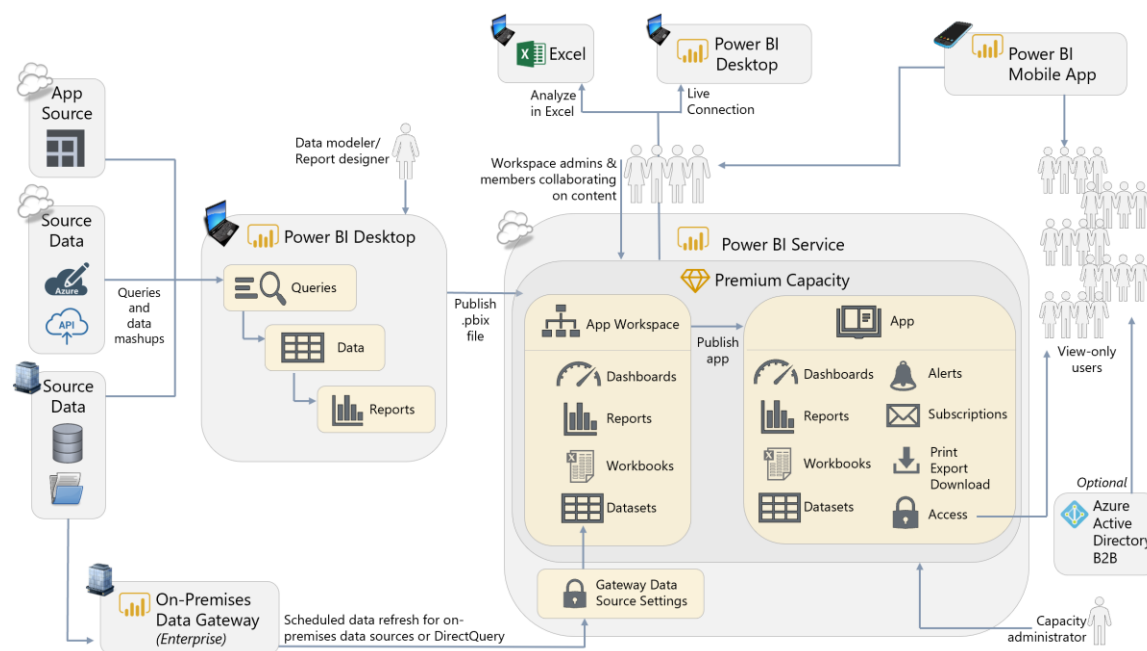
Above architecture consists of similar level of components in addition to thereports can be deployed to the App workspace, where the user can interact and collaborate to work on the reports.

PBI App can be created in the APP workspace. PBI APP is a pre-configured set of components where as dataset, reports, and dashboards together.

Enterprise collaboration and distribution involves distributing the large content to the huge audience through PBI APP, the users can be restricted to access the data. Only authorized people can access the data.

5. Centralized Corporate Reporting and Data Warehousing

© 2012 IJFANS. All Rights Reserved,
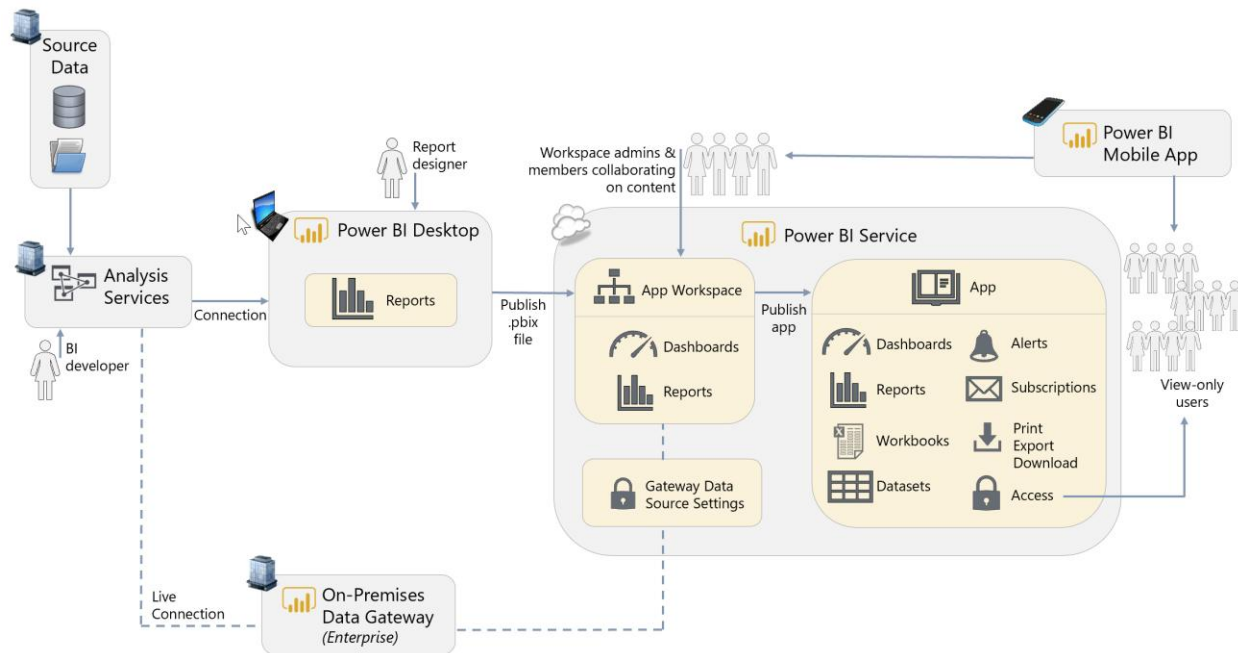
In addition to the above architecture there are few additional components has been added.

➔ Alerts Mechanism
➔ Delivering the reports notifications to the end users through subscriptions
➔ Print and export downloaded format.
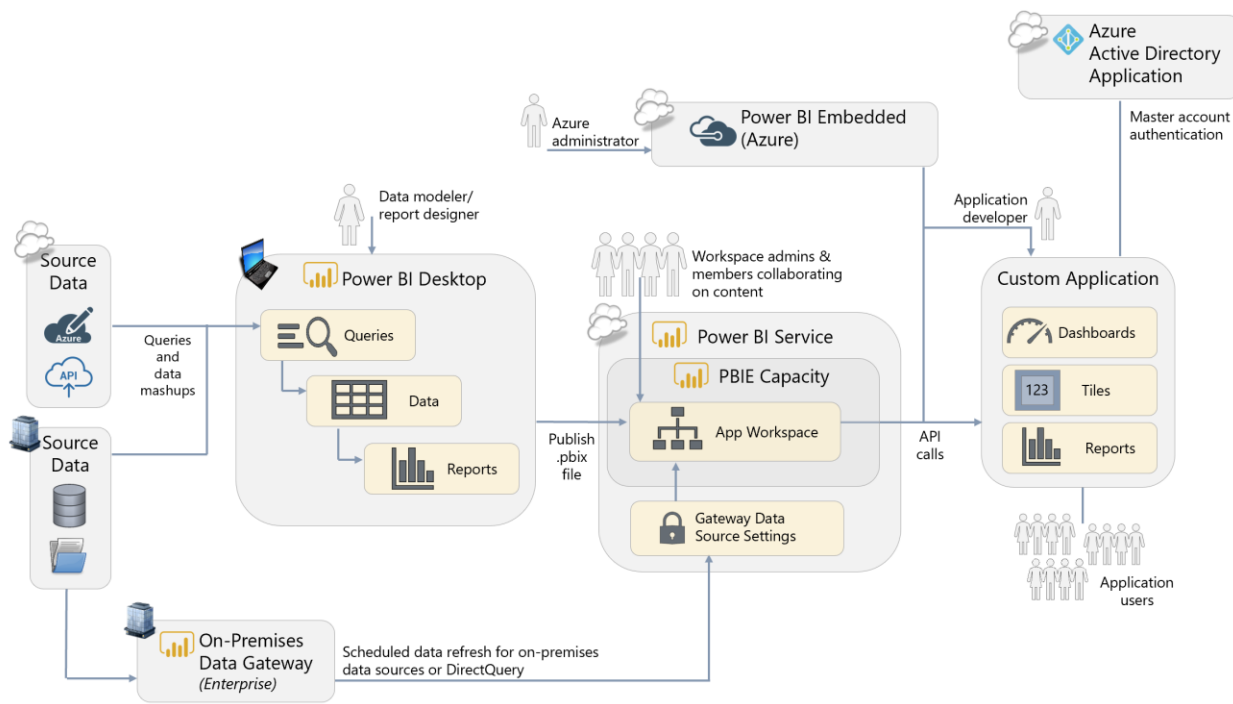➔ Authorized access to privileged users.

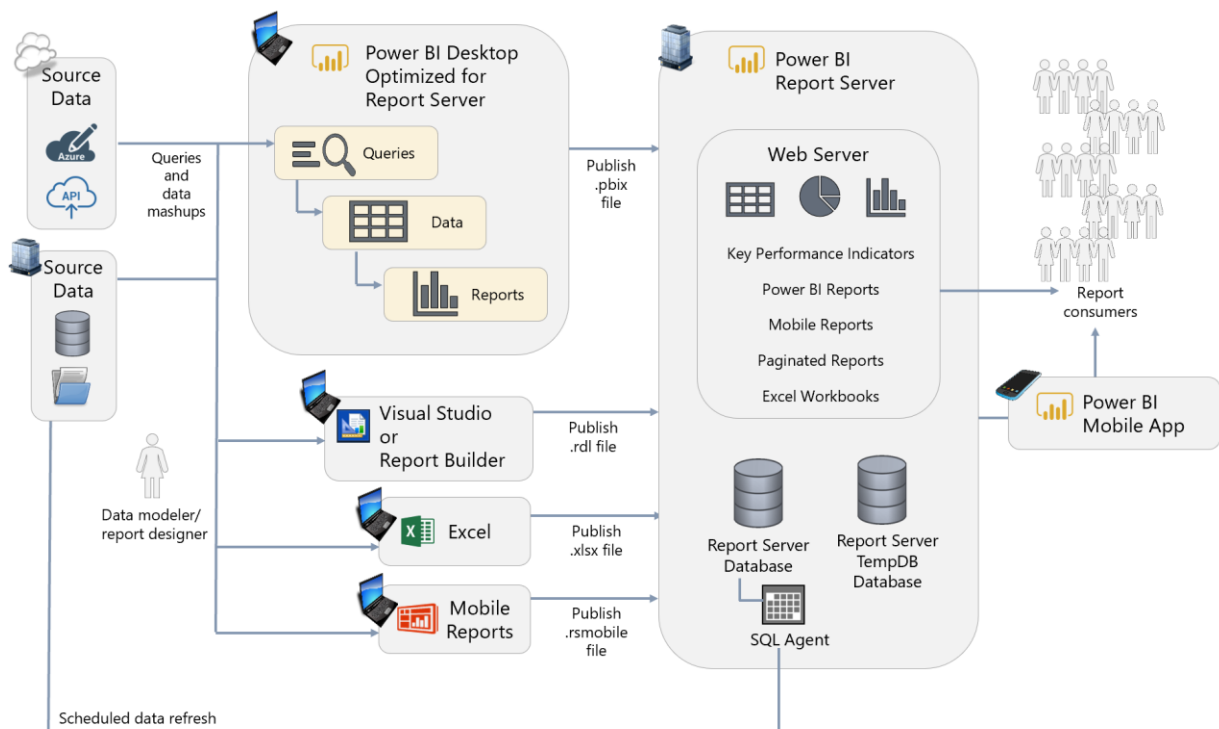6. Prototyping Activities and Sharing
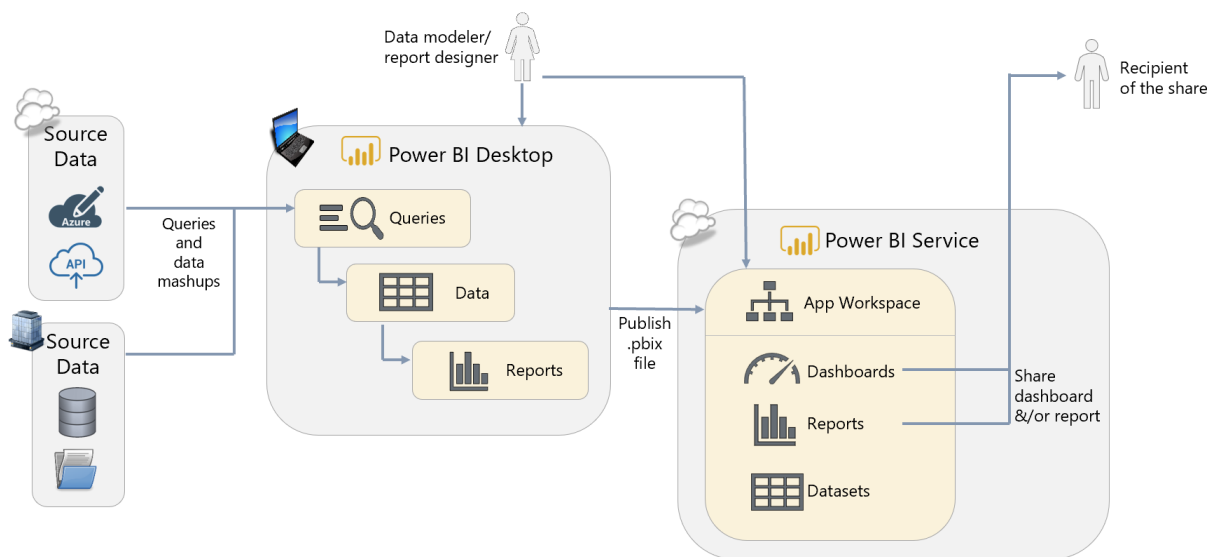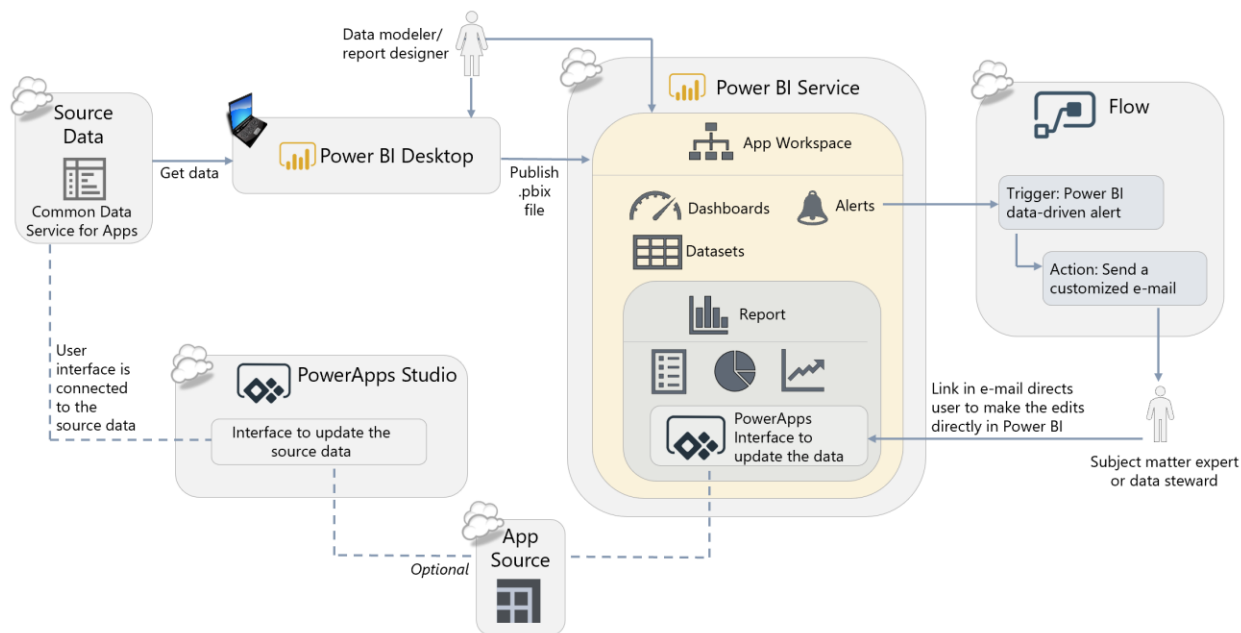
7. Organizational Embedded Analytics
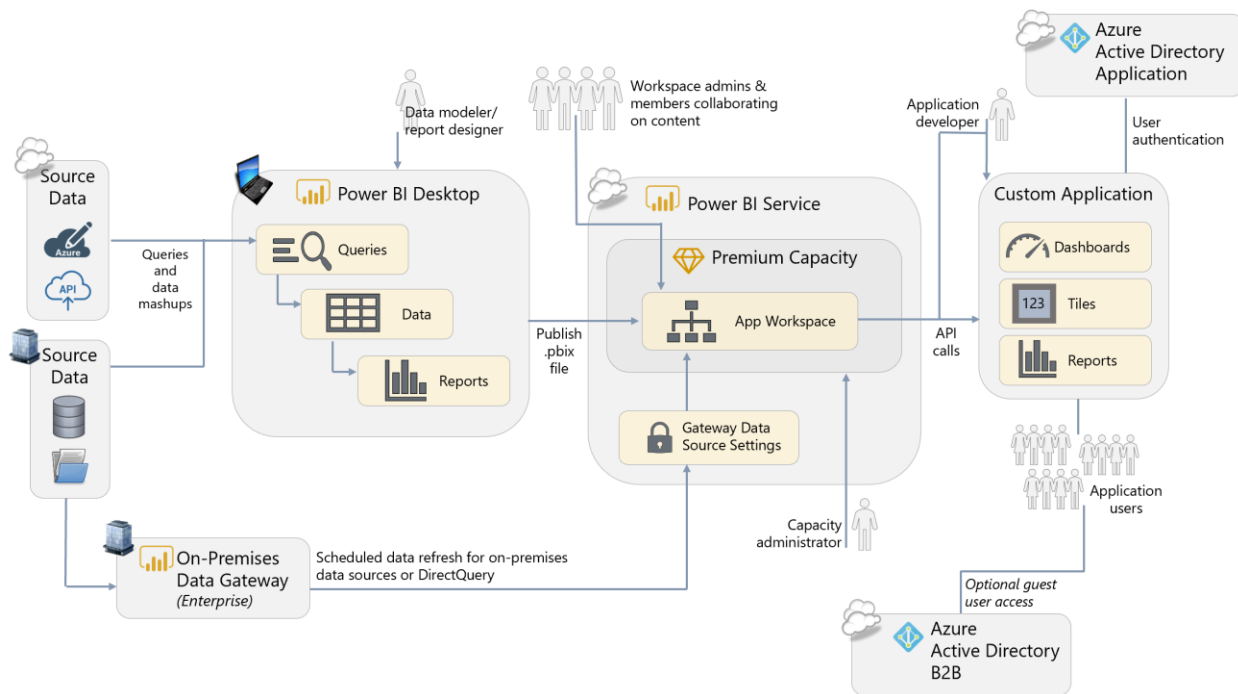
8. External Embedded Analytics

9. On-Premises (Non-Cloud) Reporting Portal

10. Workflow Automation Updating Data from Within Power BI

11. Sales Report

Sales report which comprises various sales among multiple category wise, product wise, country wise data analysis.

Materials and Methods

We applied cohort analysis as the critical method to gain insights from this research. Its usage allows efficient detection of the impact and changes in user behavior over time. The critical factors for cohorts' selection are age and social status. We consider it appropriate to add a factor such as a gender distribution in e-commerce and apply some descriptive analytics to it. According to the latest research, the involvement of women in higher education will solve the global warming problem and thus create a more eco-aware society. So, the impact of gender distribution in e-business and business information remains unexplored. It can provide a large number of insights regarding consumer preferences and understand the changing market trends.

The following methods were used for conducting the research:

- Statistical and descriptive analysis of Web users' e-commerce awareness;
- Time-series and cohort analysis;
- Correlation analysis of e-commerce aspects;
- Statistical hypothesis testing to study the difference between representatives of various e-commerce awareness levels and their perception of various e-business aspects from a customer point of view;
- Gender distribution and e-commerce awareness and loyalty levels;
- The dynamics and level of multifunctional IT solutions and business informatics usage;
- A qualitative and quantitative review of the target audience;
- Observation of the formed cohorts' e-commerce customers.

Our target audience is web users who interact with e-commerce solutions. The beneficiaries are the subjects of IT solutions implementation, research institutions, business schools, and institutes.

We used the "Survey Data on E-commerce awareness" from the Kaggle Platform [57]. It contains 204 observations with 34 features that were used for analysis. The data set was pre-processed with the following changes applied: columns renaming, data cleaning, and categorizing features. We converted some categorical features into numerical according to the rules represented in **Table 3**. The assigned numerical values are based on the question context and response sentiment: positive/negative/neutral towards e-business receive the highest/median/lowest value correspondingly. It is worth pointing out that these values depend on customer awareness and loyalty, defined in previous sections of this research. The first group of questions with possible answers "Yes", "Maybe", or "No" are primarily focused on assessing customer's awareness and more minor on loyalty. In contrast, the second and third groups with "Agree", "Neutral", or "Disagree" choice of answers consist of equal concentration on these two characteristics. For this reasons the assigned numerical values are different.

Table 3. Data conversion for e-commerce awareness and loyalty score assignment.

These features were summed up and the values were used as a score indicating engagement into e-commerce and the awareness and loyalty levels. The maximum possible score value might be 12, whereas the minimum can be −8. We applied data standardization to the score values to make them consistent and keep the outliers but reduce their impact on the calculated statistics. The mean and standard deviation values have changed to 0 and 1, respectively.

The data pre-processing part allowed us to simplify the exploratory data analysis and assign a score to each respondent based on the answers provided.

Results

The research analysis is based on data from e-commerce users' surveys and online purchase history during the COVID-19 crisis. We have provided a specific data set description and references further on in this section. We conducted a quantitative analysis on a sample of e-commerce users from different locations and with a wide range of interests.

Descriptive statistics in this paper is generated using data wrangling and visualization techniques. The fetched information highlights differences among various user groups based on their e-commerce awareness and loyalty scores. The obtained insights are helpful for monitoring and evaluating consumers' e-businesses perception, the likeliness of interacting with a platform, satisfaction rate. Moreover, this analysis leads to the development of multiple types of open innovation dynamics, which is critical for sustainability in e-commerce and e-business [58].

Engineering is required to solve various problems of open innovation, mainly the channels for knowledge sharing [59]. We recommend correlating the analyzed factors with business and performance metrics to evaluate and enhance customers' shopping experience.

The analysis contributes to an overall improvement in consumers' understanding, their perception of different e-business aspects, and helps adjust questions in further surveys on related topics.

4.1. An Introduction to Exploratory Data Analysis

Our data contain both male (36%) and female (64%) responses. The age distribution is shown in **Figure 5**, where 85% has the population within the 15–30 age range, and 11% and 4% of the data are 31–45 and 46–60 age ranges, correspondingly.
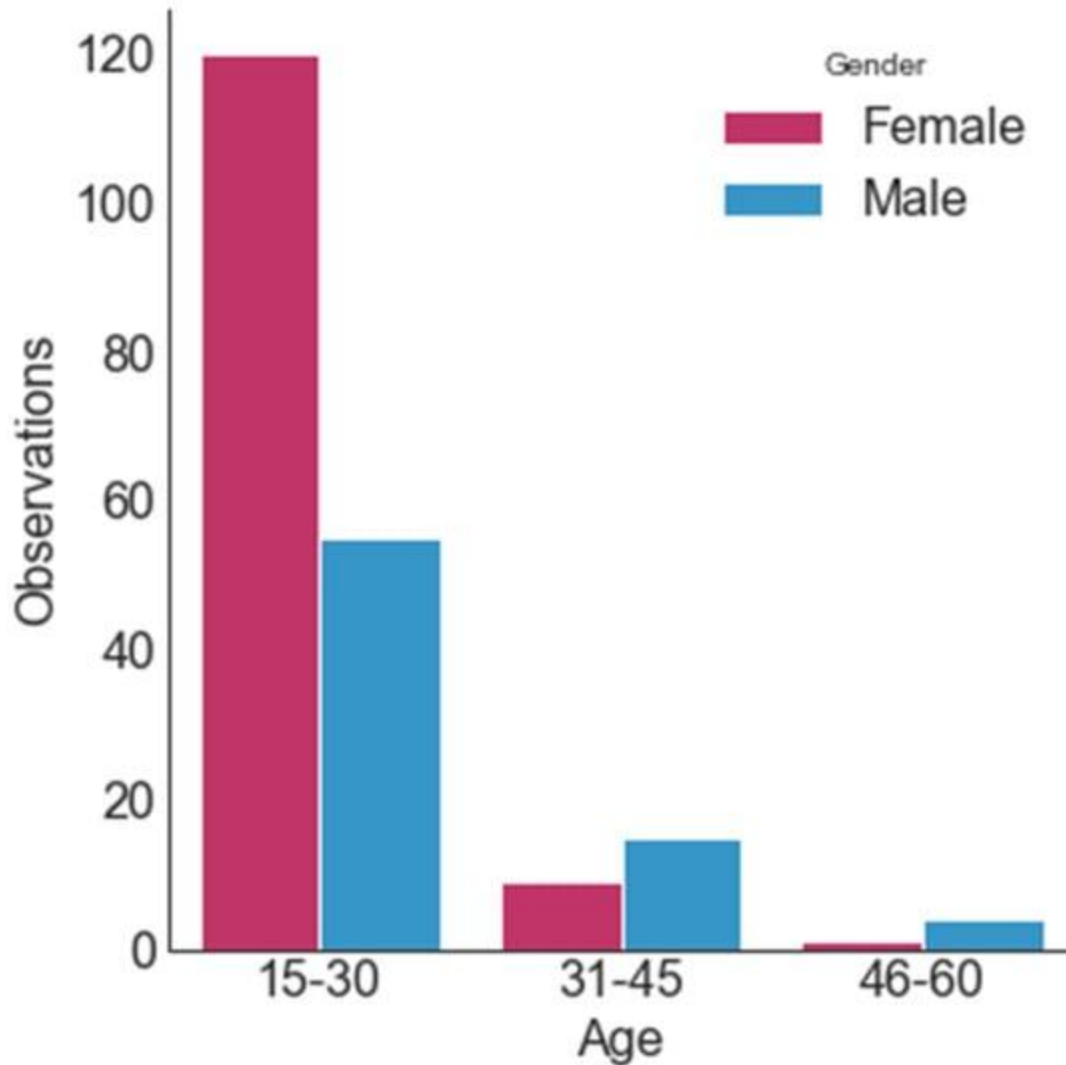
Figure 5. Sample age distribution by gender.

A histogram for the standardized score displayed in **Figure 6** follows the Extreme Value Type I Distribution pattern, also known as the standard Gumbel distribution (minimum case). Few outliers are observed, which indicates poor e-commerce awareness levels.
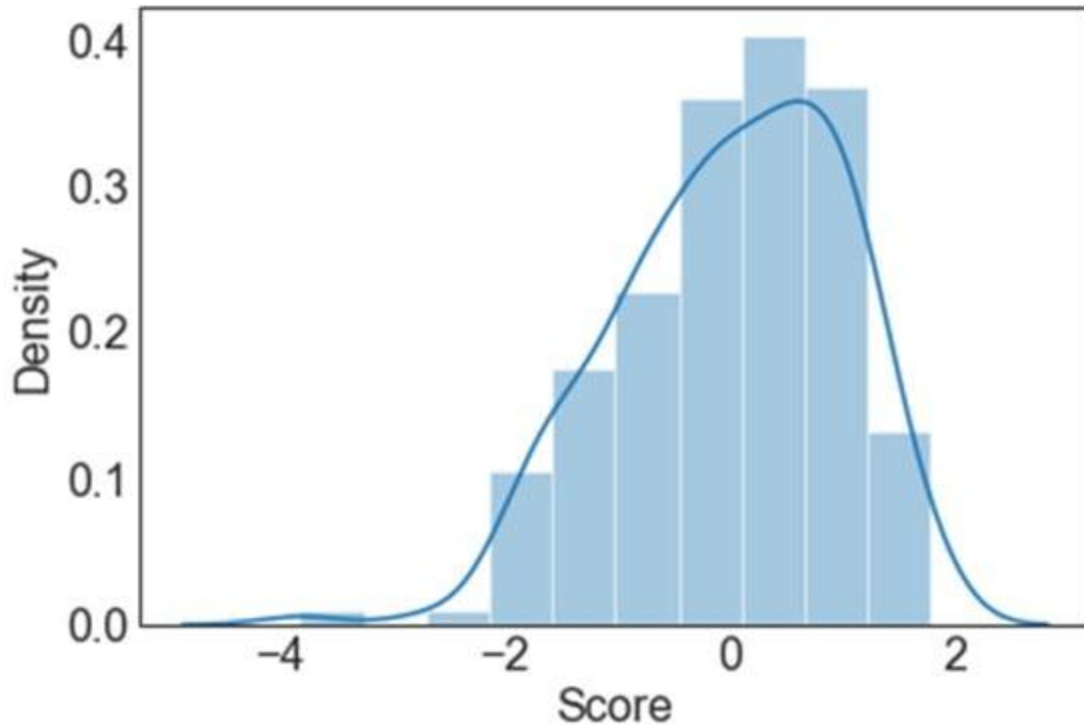
Figure 6. E-commerce awareness and loyalty score distributions.

We found out that 46% of the data points are less than 0; the excessive number of observations lie in the interval between 0 and 1 (37%); and the right tail of the data has the proportion of 17% with a maximum value of 1.77.

So, we infer that approximately 54% of the observations have a greater than average e-commerce awareness and loyalty score.

Further analysis revealed that male representative tends to have a higher score compared to female. This is shown in **Figure 7** that tells the overall distribution. Few outliers are highlighted as well. The most common values are in the range from 0 to 1.
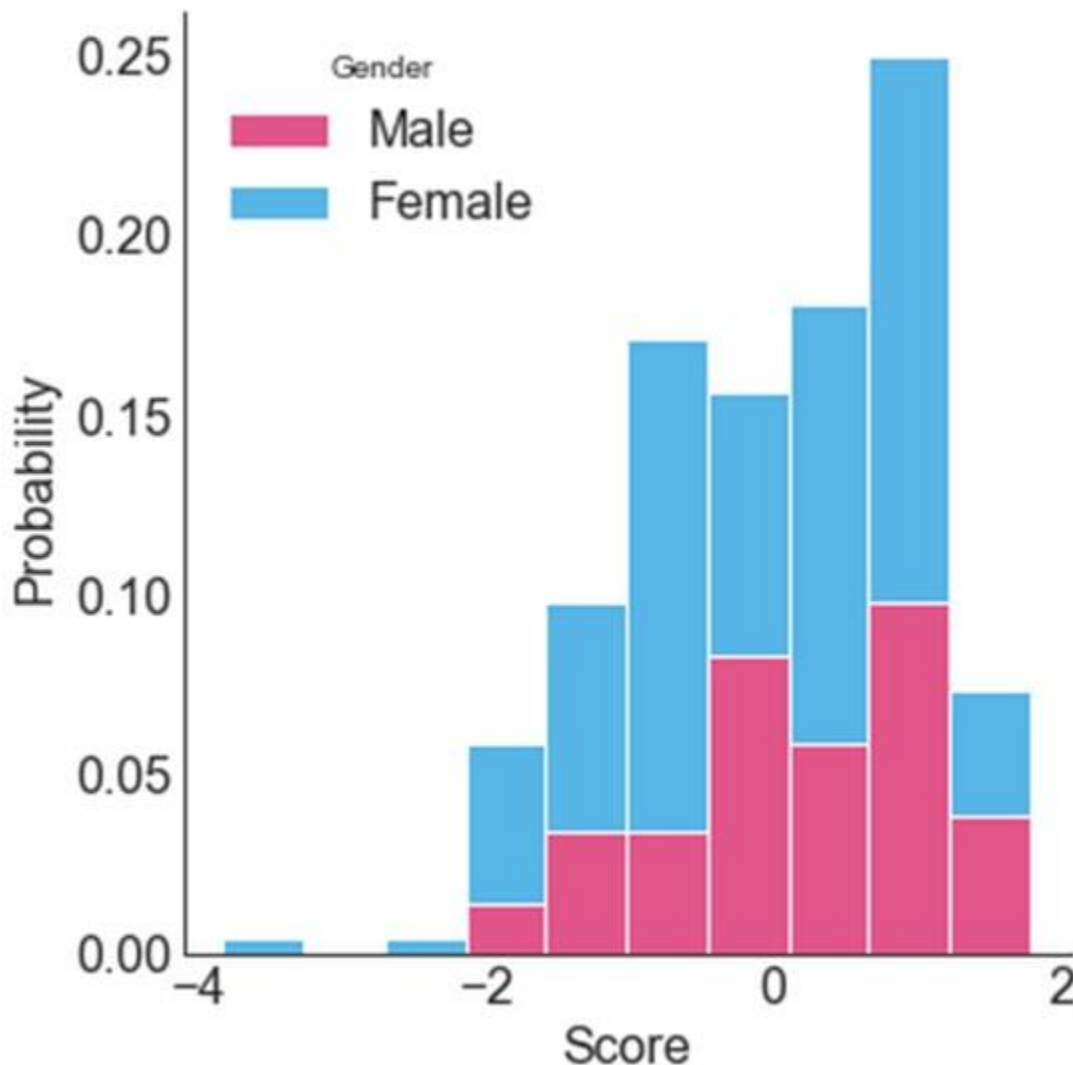
Figure 7. E-commerce awareness and loyalty score distributions.

The statistics are represented in <u>Table 4</u>. It is crucial to note that males' mean and median scores are equal, but not for females, which is caused by outliers. The percentiles are also considerably higher for the male representatives. These insights indicate that females tend to trust e-commerce less than males, thus being careful when selecting a platform for shopping.

Table 4. Descriptive statistics of the standardized e-commerce awareness and loyalty level score.

4.2. Research of E-Commerce Aspects Customer Experience and Importance

The participants were asked to rate the following e-commerce aspects from a customer perspective on a scale from 1 to 5, where 1 is the lowest score, and five the largest: affordability, usability, a time-saving, wide range of products (wide-ranging), and reliability.

The participants could assign the same score to one or multiple aspects. Because some aspects of the same respondent might have identical scores, we accounted for each aspect during our analysis. In rare cases, all five aspects could have the same score. So, we calculated the mean of them, and if it was more significant than our threshold (3), then it was considered as a positive aspect and assigned a label "max"; otherwise, it was a negative one with a label "min."

In **Figure 8**, we visualized the probability of occurrence of a particular aspect with the "min" (negative) and "max" (positive) labels.
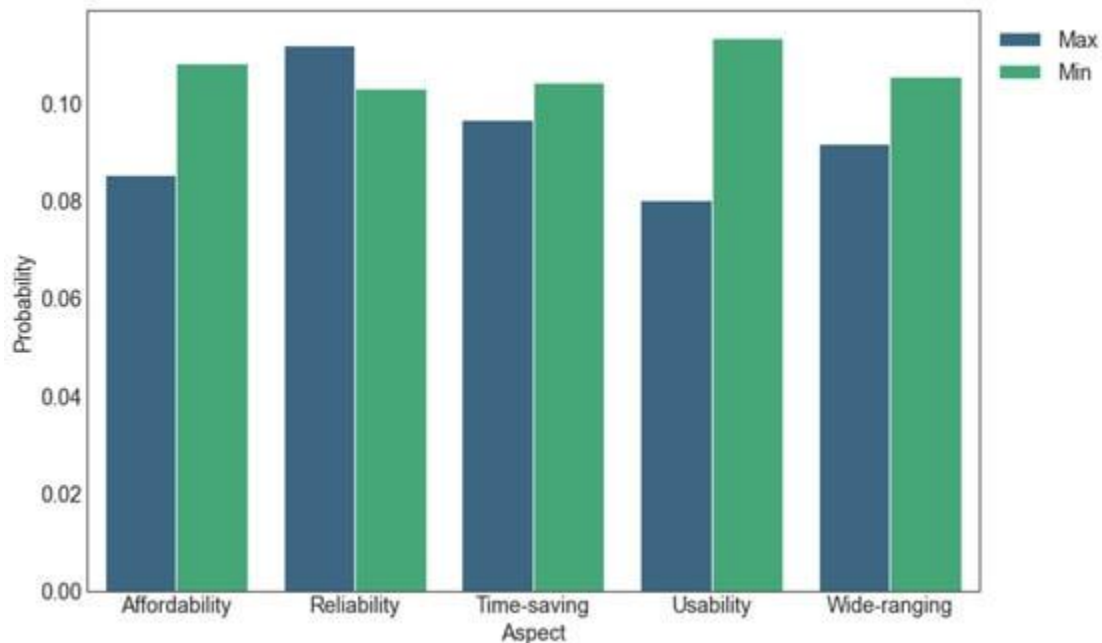

Figure 8. E-commerce aspects importance from a customer point of view.

The data contain more low-scoring aspects (52.9%), considering them as e-commerce drawbacks rather than high-scoring (47.1%). The highest probability of being negative has the usability component. It has the least likelihood in the top part as well.

Reliability is the key positive aspect and the only one that received higher than low scores.

In the study, we covered the correlation of crucial aspects depending on the calculated e-commerce awareness scores. This review made it possible to model how a particular user perceives e-commerce, and to determine the correlation between the e-business awareness level and the critical customer factors.

To avoid bias and extreme values impact, the mean (a) and median (b) values were used for the visualization in **Figure 9**. According to this graph, people with a high score are pretty optimistic about its usability and that it is wide-ranging. However, the affordability and partially time-saving aspects have received low scores from these users.
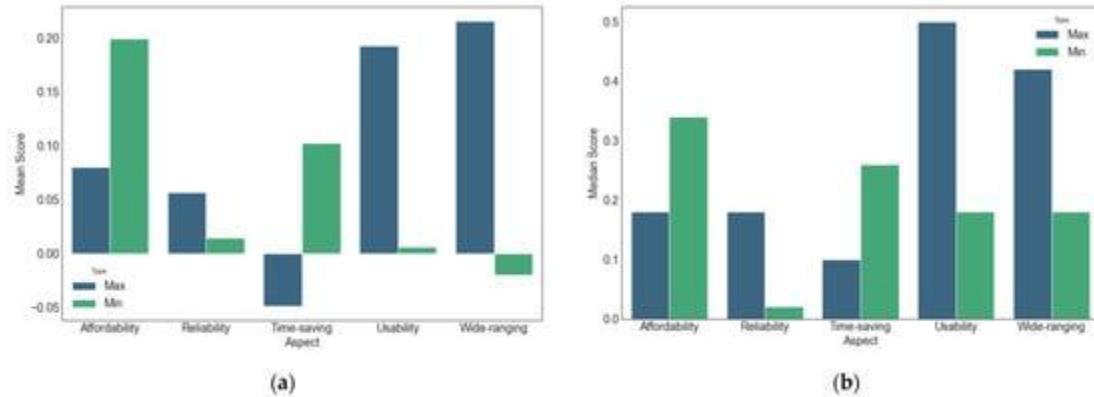
Figure 9. (a) Correlation of user aspects importance with the mean e-commerce awareness score; **(b)** correlation between user aspects importance with the median e-commerce awareness score.

On the other hand, users with a lower than the average score consider the e-commerce process time-saving (utterly opposite to the first group of users). They tend to evaluate it with a value bigger than three. They do not suppose e-commerce to be wide-ranging.

The third type of user with an average e-commerce awareness level positively evaluates affordability, whereas usability is not an advantage. It is instead a part of e-commerce that needs improvement.

A study of the purchase frequency on e-commerce platforms gave predictable insights visualized in **Figure 10**. The more frequently a person buys a particular product or uses specific services, the greater awareness in this area. The type of users who never use e-commerce tend to have a lower awareness level. Weekly shopping frequency representatives have the highest median score as well as higher quartiles compared to other categories. The users who use e-commerce daily have a minor representation and a pretty decent standardized score, mainly between 0 and 1, with two outliers.
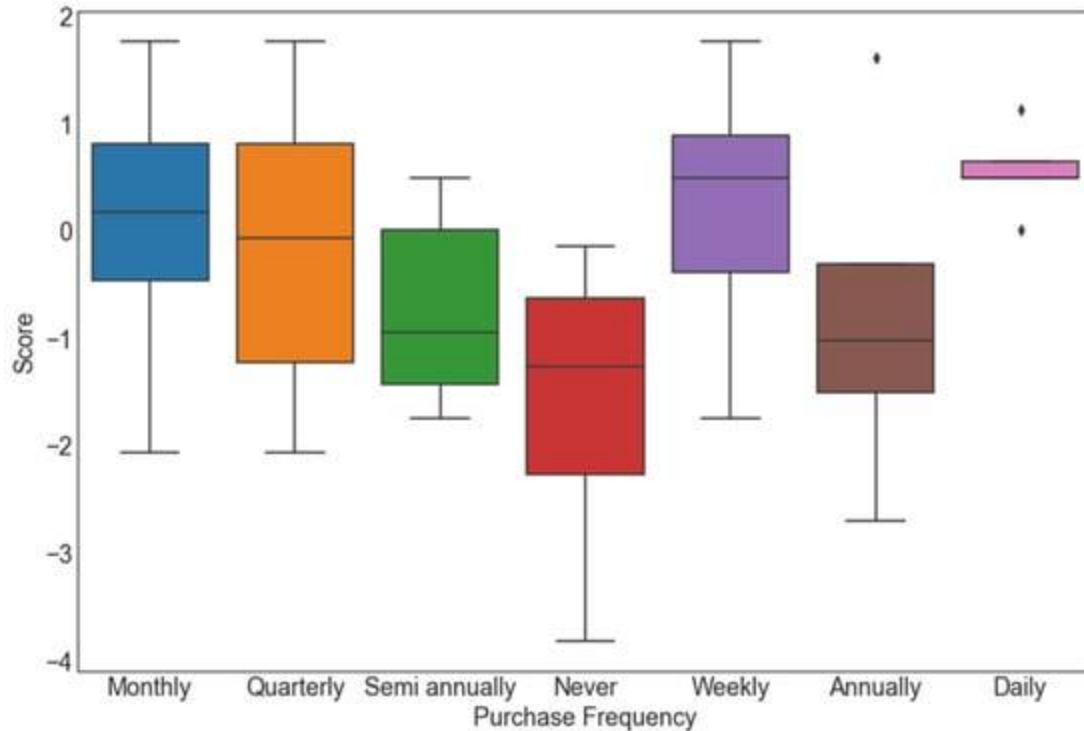
Figure 10. Standardized e-commerce awareness score distribution by purchase frequency.

4.3. Reliability Aspect Hypothesis Test and Deep Dive

Reliability is both a positive and negative aspect for users across this awareness level. However, a trend confirms that the higher the user's awareness score, the more inclined this user is to evaluate the platform and the industry as reliable.

A hypothesis test was conducted during the research to determine whether there is a statistical difference between two consumers groups for their reliability aspect perception depending on the e-commerce awareness score. The first group considers e-commerce to be reliable, while the second as unreliable. The sample means the difference in the standardized score is 0.042. The H0 indicated no difference in the populations' awareness score, whereas the Ha indicates the opposite. As it turned out, we failed to reject the H0 hypothesis with a significance level of 5%, meaning that there is no statistical difference between the two populations who have an unequal perception of reliability.

In order to provide better consumer understanding, thorough data analysis was conducted for those users among whom reliability belongs to the most positive or negative aspects of the e-commerce platform. It should be noted that 52% of respondents positively assess reliability. As it turned out, users aged 31–60 with above-average e-commerce awareness evaluate reliability as a competitive advantage of e-business. This is shown in **Figure 11**.

Research Paper   © 2012 IJFANS. All Rights Reserved,
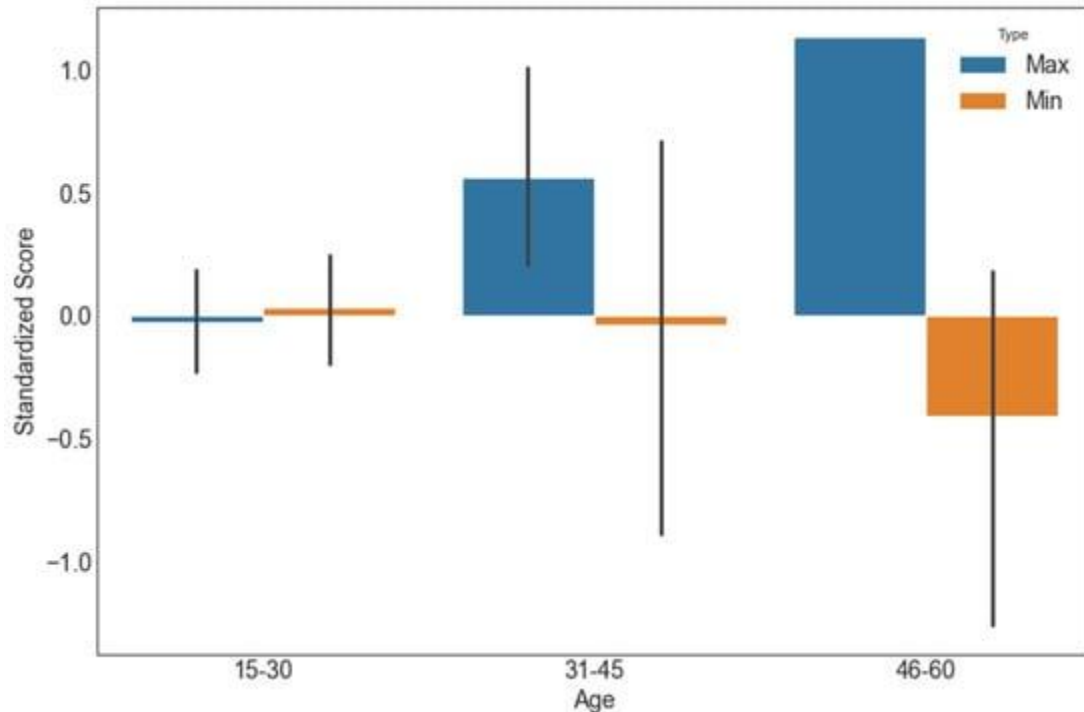

Figure 11. Standardized e-commerce awareness score distribution by age range for Reliability aspect.

The exact correlation is seen in e-commerce aware users of the "Businessperson" occupation. At the same time, users aged 15–30 with a score close to the mean think of it as a drawback. Respondents whose e-commerce purchase frequency is shallow or does not even exist do not consider e-commerce reliable.

The same applies to respondents who make purchases mainly on a local market or retail store. This reaction might be the cause for making no purchases there or making them infrequently. Above 70% of customers who mostly purchase in the electronic category think e-commerce is unreliable; only 10% of restaurant food e-commerce customers think vice versa.

A total of 65.7% of respondents whose preferred payment option is credit/debit cards, think of e-commerce reliability as a negative aspect that needs further improvement.

4.4. US E-Commerce Records Cohort Analysis

The US 2020 E-commerce records data set was used to implement the cohort analysis. The data collection happened from January to December 2020 [60]. Cohort analysis can be very useful for analyzing time-series data to find patterns of change in customers' behavior. It allows gaining a lot of valuable insights and exploring a vast range of metrics, corresponding trends. The customer's monthly retention rate was calculated as a part of the analysis. The retention rate allows measuring the number of customers a company retains over a specific period [61,62]. The cohorts were formed based on transaction date features, and each user was assigned to a specific index depending on the interaction with an e-commerce platform. The result is visualized as a heatmap shown in **Figure 12**.

Figure 12. Standardized e-commerce awareness score distribution by age range for Reliability aspect.

Cohort index 1 shows users who did their first interaction with the e-commerce platform on a specific month. If we look at cohort index 4 and the 2020-01 cohort month, we derive that only 24% of users made purchases in 2020-04 after being first introduced to the system four months earlier. This is the most significant value among all cohort indexes for this cohort month users. The highest retention rate equals 31% and is observed for the cohort month 2020-09 users at index 3. So, based on these statistics, we can answer the following business questions:

1. What is the customers' retention percentage?
2. What month has the most interactions between users and the e-commerce platform?
3. Does a specific cohort month have an impact on the retention rate? Is it related to any technical changes on the platform/promotions/business events?
4. At what index the customer retention starts increasing/decreasing?
5. What retention rate can we expect in the near future based on the current trend?

Solutions to these and other questions can be found using cohort analysis. From the chart above, we can see that the retention rage does not go above 31% and is 18% on average. We can clearly claim that only between 7% and 31% of customers were retained for the analyzed period. However, it does not mean that the same users interacted with the platform for the whole period. We can assume that a specific cohort sample interacted at index i. Another sample that might or might not contain previous sample users did its interaction at index i + n, and so forth.

5. Discussion

Business informatics is a very rapidly developing discipline with a lot of methods and applications to be studied. There are several educational programs and initiatives in business informatics in Europe to provide the labor market with resources capable of managing data in the business environment.

According to statistics from the online service Keystone, there are 54 bachelor's and 39 master's degree programs in a set of business informatics disciplines. The study duration in these programs usually is from 6 to 8 semesters and covers the following aspects: computer analysis, data management, IT project management, computer systems, and software design. With the technological development, emphasis should also be placed on Big Data Analysis, Finance Machine Learning, and Data Science to be included in these courses. The number of such programs only increases, confirming the demand for business informatics, IT data, and knowledge management professionals in Europe and globally. Exchange of experience with business informatics specialists through the Web is necessary to conduct advanced research and implement the experience and results in the global academic environment. In particular, to create appropriate educational programs, methodological support for high-quality business informatics teaching, and organize corresponding international scientific and practical conferences. We have provided insights into the current research and development of cohort analysis usage in e-commerce and, essentially, business informatics. The emphasis is on the interpretability of the results that this technology delivers and the need to correlate various types of analysis with descriptive statistics. From a user perspective, the study of e-commerce aspects reveals valuable knowledge on what satisfies and frustrates users during online shopping. It can allow identifying users that need particular focus from the e-business representatives to increase the KPIs as well as e-business aspects that need improvement. This is especially critical during crises where trust is required for long-term business relationships with clients. We recommend researching implementing a strategy for efficient cohort analysis usage within specific industry representatives. This further study might explore tools, methods, and required resources regarding defined best practices.

The cohort analysis method requires the same user to have at least two actions completed over a period in order to form the cohort and provide knowledge. Nevertheless, the real-world limitations could complicate this process, and data might not be available. For instance, the user has not yet made more than one transaction or has temporarily stopped interacting with the system. Thus, this person cannot become a part of a cohort. We urge researching the possibility of cohort formation based on multiple users with similarities and representing them as a single entity. As an outcome, a cohort will contain N entities, where each of these entities contains M users with shared characteristics. Hence, the likeliness of collecting enough data about a specific entity is higher than when the entity consists of only a single person.

Open innovation in e-commerce can help establish an industry-validated framework for conducting cohort analysis and increase the interpretability of the obtained insights. From a long-term perspective, this will allow the presenting of high-quality research on cohort analysis and e-business. Recent studies shows that open innovation can lead to sustainability in the fourth industrial revolution. Yun J.J and others developed a quadruple-helix model to understand its dynamics in various areas [63]. Studies show that open innovation and technologies like artificial intelligence can help overcome post-COVID-19 challenges across organizations [64,65]. This fact is vital for e-commerce and other digital industries that have faced the fight against the pandemic period.

It is critical to research how software performance and infrastructure availability can engage new users and achieve a high retention rate. Thorough research on user engagement and its correlation with e-business service metrics has shown various methods for assessing system performance and customer satisfaction. This research can be extended with the cohort analysis application on the e-commerce data set, and new knowledge can be mined. We consider it essential to study the various aspects of e-business services from a customer perspective, mainly the reliability since our analysis revealed that different user categories perceive it differently. Further work on improving this experience needs to be undertaken.

Research Paper

Conclusions

To increase the level of data-drive business awareness, e-commerce representatives should enhance their business informatics usage; cohort analysis is one of these efficient methods that allow organizations to improve it. We conclude that in this way, these actors will make their business more observable than previously, detect gaps in the technological aspects, and accelerate the path of progress. As a result, e-commerce will move at the new pace of evolution of its representatives, develop new products and services, and meet the expectations of a new type of consumer. Adequate usage of the cohort analysis method will provide institutions with the necessary information resources to form an intelligent enterprise and a framework for crisis management. As an example, when considering the COVID-19 crisis, lack of specific practice negatively affected the dynamics of growth and turnover. Those representatives who already successfully used business informatics obtained the following results: the presence of multifunctional management information systems, quality business solutions, and efficient resources allocation. Once the business informatics best practices are followed, the cost and efficiency of e-commerce will be the most optimal. Thus, the study of COVID-19 impacts on the dynamics of business informatics advancement with the web users' cohort analysis will increase e-business efficiency, identify trends in products and services, factors influencing current results, and model the progress of the events above in the future. The derived insights on e-commerce awareness could bring value to business persons. Our findings show a clear difference in the experience of customers on different understanding and loyalty levels. People with an above-average e-commerce awareness score consider e-commerce not as affordable and time-saving as other people with lower levels assume. The difference between the judgments of various aspects, mainly the wide range of products, in the e-commerce users' behavior depends on the assigned score. For instance, respondents with high scores consider this an advantage, thus confirming the presence of a wide-ranging aspect in the e-commerce platform, whereas less aware participants think the opposite. These and other insights obtained will help make appropriate business decisions on the following items:

- Technical and/or business aspects that need improvements;
- Customer perception of e-business;
- User satisfaction/dissatisfaction;
- Observability of the customer journey and infrastructure performance.

Moreover, an e-business might discover what new data it needs to collect to mine extra knowledge, understand whether the industry standards are met, and focus on the next steps. On the other hand, if more focus is on a particular e-commerce aspect than others or the e-business, there is a certain probability of losing specific groups of users. So, a balanced attitude towards aspects improvements should be taken into account.

However, implementing the mentioned above cohort analysis usage might require resources. Because of this, we recommend doing so successively with continuous evaluation and optimization. The results need to be adequately interpreted, and sometimes, external consultants need to be hired to guide towards the next steps. If we take the whole population of customers in the cohort analysis, there is less impact on the types of people participating in it. Moreover, in some cases, the researchers cannot identify to what cohort-specific user belongs or the reasons for being a specific cohort part. It is worth mentioning that cohort analysis might provide value in the long run. In our study, we used the transaction date characteristic as a common one for grouping users into cohorts. The date features are used across various domains and tools as well. We recommend specifying a more specific characteristic that describes the group or multiple features to form the cohorts. This approach will increase the similarity among users in a cohort and allow precise knowledge mining on them. It might include products bought during a period, interests and acquisition date, location, shopping frequency, e-

Research Paper   © 2012 IJFANS. All Rights Reserved,

business aspects perception, and awareness. The researchers will benefit from this because they will have more control over the sampling process than when there is little transparency in cohorts' formation. Characteristics can be selected based on statistical analysis and updated if changes in the real-world data occur. Big data is very promising in terms of cohort analysis usage in e-commerce. Its application in this area can reveal insight and allow efficient cohort formation. Mainly, we can use Big Data to sample users for the cohort analysis, thus having control over the data collection process. Nevertheless, if the business focuses on a short-term goal and needs results immediately, or there are not enough resources and time to execute the cohort analysis, other methods might become more valuable than this one. Further empirical analysis on industry and research-proven cases need to be accomplished. This method in e-commerce requires thorough explorations to deliver valuable results and prove expectations.