National College of Ireland

Project Submission Sheet – 2020/2021

Student Name: Heena Chopra, Jaswinder Singh, Belal Masarrat Farooqui, Ruchita Dattatray Patil

Student ID: x19205309, x19219997, x19201036, x19197411

Programme: MSc. in Data Analytics   **Year**: 2020

Module: Business Intelligence and Business Analytics

Lecturer: Mr. Vikas Sahni

Submission Due 20 December 2020
Date:

Project Title: Final Project CA2 (Implementation Report)

Word Count: 3611

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature: Heena Chopra, Jaswinder Singh, Belal Farooqui, Ruchita Patil

Date: 20 December 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1.   Please attach a completed copy of this sheet to each project (including multiple copies).
2.   Projects should be submitted to your Programme Coordinator.
3.   **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
4.   You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**
5.   All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

National College of Ireland

Continuous Assessment - II

BUSINESS INTELLIGENCE AND BUSINESS ANALYTICS

IMPLEMENTATION REPORT

Jaswinder Singh
(x19219997)

Heena Chopra
(x19205309)

Belal Masarrat Farooqui
(x19201036)

Ruchita Dattatray Patil
(x19197411)

Group - 5(Cohort 'B')

Arecibo
Pvt. Ltd

Contents

List of Figures

1 Balanced Scorecard

In the project, we believe it is vital to use a *Balanced Scorecard* also known as **BSC** because this is the best framework available to manage the company's reporting. A Balanced Scorecard is a strategic performance management framework that was developed in 1992 by Dr David Norton and Kaplan [1], which uses four unique perspectives of strategy: Customers, Financial, Internal Business Process and Learn and growth. Each feature relies on the following requirements:

- **Objectives:** High-level organizational goals.

- **Measures:** Accommodates in understanding if the company is accomplishing the objectives strategically.

- **Initiatives:** These are key action programs, developed to achieve the company's objective.

- **Action items:** Tasks are delegated to a small team to achieve key initiatives.

A well-known suggestion from veterans of Data Analytics is that if the company's strategy has three objectives, they will succeed in all three. If they have four to ten objectives, then they might succeed in one or two, however, if it contains more than ten objectives, then they will most likely succeed in none.

Our company, Arecibo Pvt. Ltd., is primarily concerned with operational efficiency, which can be achieved by increasing its product output, and by decreasing the production costs. Additionally, this can be further improved, by implementing a reasonable market price on goods, ensuring their products are easily and readily accessible to customers and making sure the right assortment of goods is being manufactured.

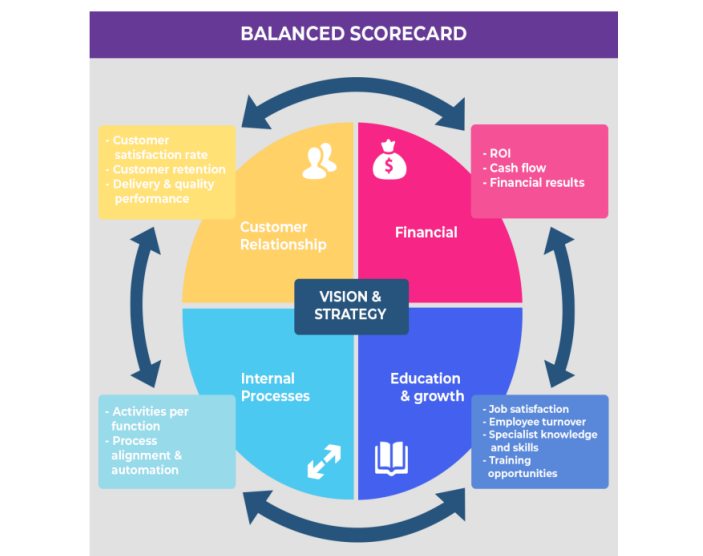Figure 1: The four prespectives used in the balanced scorecard

Figure 2: Balanced Scorecard for the Arecibo Pvt. Ltd.

Figure 3: Startegies for the Arecibo Pvt. Ltd.

2 Solution Development Process

2.1 Data Generation and Cleaning

The first and most important step in the solution development process of a Business Intelligence and Analytics system is the generation of the data. After doing a thorough background search on the companies in the DIY industry, relevant entities and their attributes were chosen for creating the data for our company. The random data generator *Mockaroo* was chosen for the data generation. The data was then cleaned and prepared for analysis using various libraries in *Python* and also *Microsoft Excel*.

2.1.1 Data Generation using Mockaroo

After gathering some information and doing a thorough background search on some companies in the DIY industry, appropriate datasets were mocked using the online random data generator tool *Mockaroo*. Different entities were mocked with appropriate attributes. The list of customers was generated with attributes such as Name, Email Address, City, Phone number and different unique IDs like *Lead ID*, *Opportunity ID*, *Account ID*, etc.

After the generation process was completed, the data was cleaned and pre-processed using various Python libraries like *Pandas*, *Numpy*, etc.

Figure 4: Demonstration for the generation of Leads schema in Mockaroo

Figure 5: Data schema for the year 2017-18. Similar process was followed for the other two years as well

2.1.2 Data Cleaning and Preparation using Python

After the datasets for all the years were generated, they were saved into different data frames in python and cleaned separately one by one. The data preparation involved the following tasks:

- Assigning of unique product IDs to each of the products

- Joining of different schema by the foreign keys.

- Keeping the closing and opening dates of leads and opportunities at least 7 days apart. The important thing to be noted here is that since mockaroo generates a random sets of dates within the specified range but does not fulfils the above condition, this task was accomplished the using the *pandas* library in python.

- Assigning of Ratings(Warm/Hot/Cold) and the probabilities to the engaged customers(opportunities).

- Assigning of the description(Won/Lost/Open) to each of the engaged customers based on the ratings and probabilities.

- Assigning of the sales price to each of the products and then calculating the revenue accordingly. The revenue column was added using various conditions in Python that were observed in the trends from sales hub.

In totality, 15 tables(5 for each year) were generated using the mockaroo and the above procedure was followed for each of the years.

2.2 Building Cloud Based Database - Amazon Web Services RDS

When any emerging organization starts and begins to expand it's business, it is faced with a lot of challenges. One of the major challenges faced by the companies is the handling of its database efficiently. As more and more customers are added, it becomes harder to manage the enormous data being generated every second. Therefore it is very crucial for a company too manage and handle its data effectively. A cloud based database provides a stable, long term and very cost effective solution to such a problem. It has advantages like flexibility for data storage, accessing of services from anywhere in the world, low maintenance cost, increased encryption for the data. A cloud based database has the same features as of any traditional database with the added benefits and robustness of cloud computing. There are many platforms that offer cloud based database services at reasonable prices to the users like Amazon Web Services(AWS), Microsoft Azure, IBM Cloud, etc. Therefore keeping all the above things in the mind, our solution development process also involves migrating the database to a cloud based service. For this, we make use of the Amazon's *AWS(Amazon Web Services) Relational Database System(RDS)*. In the RDS service, we use the postgres as our database management system. After the setting up of our company's database on the cloud, we then load the data into *Power BI* for visualization. The whole process of data migration is explained in the next two sections.

(**Note:** *The reason AWS was chosen for the purpose of the project is because it provides a ton of cloud based services to its users and its subscription was provided to the students of Data Analytics by the school of computing to use for free*)

2.2.1 Moving Data to Cloud(AWS)

The database was moved to the cloud using open source development platform for postgreSQL called *pgAdmin*. The following steps were followed while moving the database to AWS.

- **Step 1:** For creating an instance on the AWS, first the user login was created on the Amazon AWS (using the credentials provided by the college).

- **Step 2:** After that, the RDS instance was created on AWS by following the instructions given in the AWS Documentation (Refer to Figure 6 for the details of the instance).

- **Step 3:** The engine(database management system) selected in *Step 2* was *Postgres*. The end points and ports were selected carefully so as to enable the connection with pgAdmin (Refer to figure 7).

- **Step 4:** After that the pgAdmin(version 4) was installed into the system from their official webpage. The master password was chosen for later use.

- **Step 5:** After the installation of pgAdmin, the connection was established between the RDS instance created on AWS and the pgAdmin 4 using the endpoints (Refer to figure 8).

- **Step 6:** Once the connection between the pgAdmin and AWS was successfully established, the tables in our database were created in pgAdmin console using the query tool (Refer to Figure 9)

- **Step 7:** After the tables were created, they were loaded with data which was present in the csv files(generated from mockaroo) (refer to Figure 10)

Figure 6: Description of the AWS instance created

Figure 7: Endpoints of the instance

Figure 8: Connecting AWS to pgAdmin 4

Figure 9: Creation of tables in pgAdmin

Figure 10: Tables loaded with data generated from Mockaroo

2.2.2 Moving Data from Cloud(AWS) to Power BI

After the data was successfully migrated to AWS RDS postgres server, the following steps were followed for moving the data to Power BI.

- **Step 1:** The AWS public key was downloaded from AWS website.

- **Step 2:** After that it was converted to a PKCS7/P7B type certificate with the help of SSL Converter (Refer to Figure 11).

- **Step 3:** The certificate was then converted to a trusted root certificate using the guide here.

- **Step 4:** After that, the postgreSQL server was chosen in the get data option in Power BI to load the data successfully (Refer to Figure 12).

Figure 11: Conversion of AWS key to PKCS7/P7B certificate using SSL converter
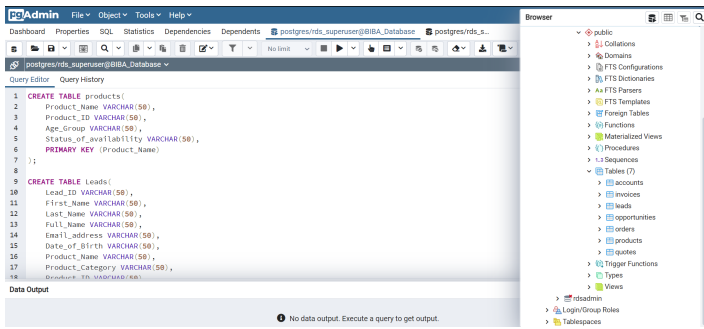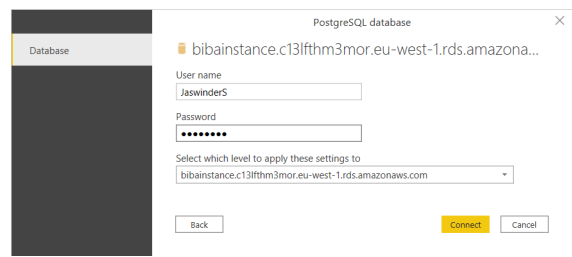

Figure 12: Loading data into Power BI from AWS RDS server

3 Implementation of the Solution

The implementation of the solution involves two aspects:

1. Integrating the customer data on a Customer Relationship Management(CRM) software. For the purposes of this project, Microsoft Dynamics 365 was used.

2. Creating interactive dashboards in Power BI to visualize various trends in the customer and sales data.

The full pictorial representation of the solution implemented is shown in the Figure 13.
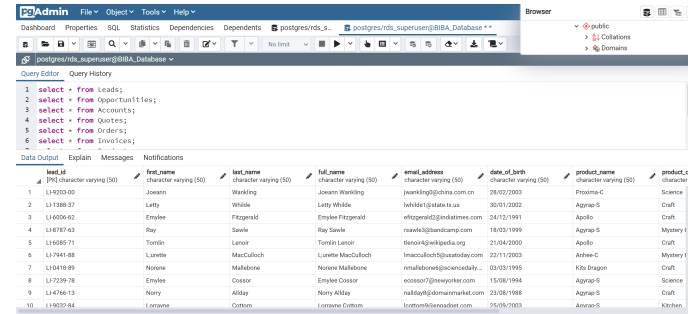

Figure 13: Pictorial representation of the implemented solution

3.1 Integrating the Data on Dynamics 365 Sales Hub

Dynamic 365 sales hub is used to build Customer Relationship Management(CRM) systems. In our project, we have used a sales hub to transform data within the company in easy manner and we can save our data for future use. The process followed is summed up in the following screenshots.

Figure 14: Step 1: We created leads to fill customer details and marked them as qualified for further process, for certain reasons if we want to dismiss the process there is a disqualified option with 4 options like lost, cannot contact, No longer interested, Cancelled.

Figure 15: Step 2: After qualified leads, it is converted into opportunity. In opportunity, there are stages like develop after that propose and close. In opportunities, we can create a product and if it is successful, we create a quote.

Figure 16: Step 3: If our quote was successful then we can select the option to create an order all good then we can create an invoice for a particular customer, but for a certain reason, our process is stopped then we can mark it as lost.

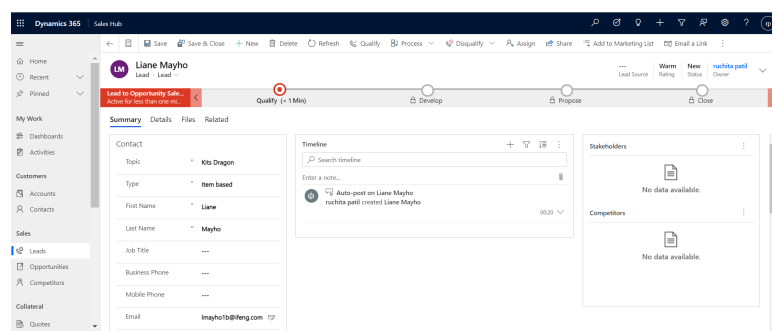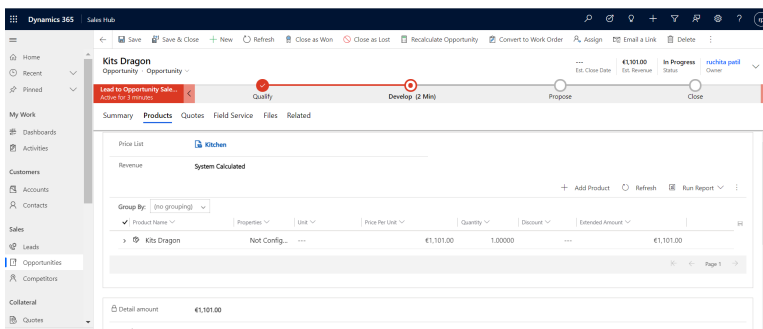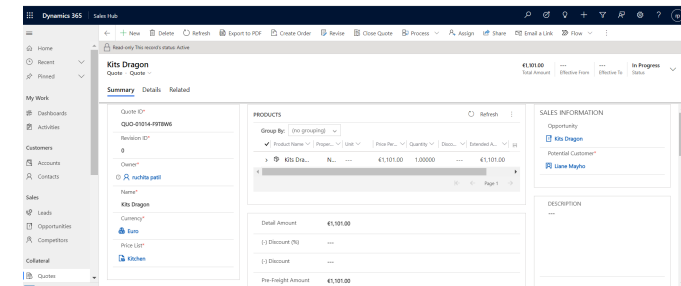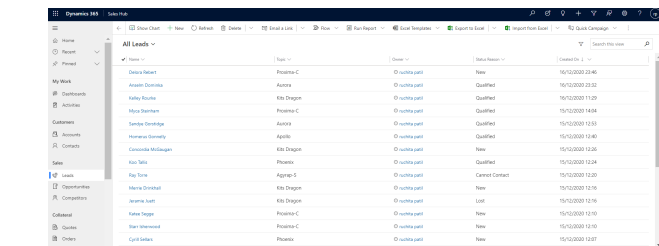Figure 17: Step 4: We can export all details into an excel sheet. The above screenshot shows all leads details. Like lead, we can export opportunity, quote, order and invoice list.
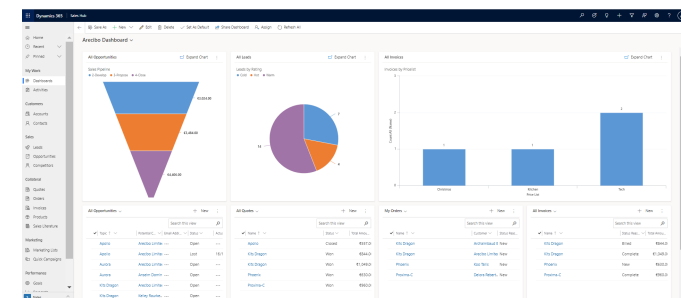

Figure 18: Step 5: Dashboard represents overall details of the supply chain. We have shown leads graph rating wise, all opportunities sales pipeline and invoices by pricelist wise. And below listed all Opportunities, Quotes, Orders and Invoices.

3.2 Dashboards in Power BI

3.2.1 Sales/Revenue Dashboard

The Sales dashboard shows the trends in sales of different products and the revenue generated over the years (refer to Figure

- The selection switch on the top enables user to view the trend by each *Product Name*, *City*, *Age Group* and *Product ID*.

- The map shows the revenue generated in the current year in different cities in Ireland. The user can also view the value of sales revenue generated in each city by hovering over the pointer to the city in the map.

- The *Revenue by City* bar plot shows the distribution of sales revenue by different cities in all the years. The Carlow county records the highest no. of sales for two consecutive years 2017-18 and 2019-20. But in the year 2020, when the solution was implemented, the trend has changed drastically.

- The *Net Revenue by Age group* pie chart shows the proportion of net sales revenue generated shared by the different age group products.

- The *Revenue By Product* bar plot shows the revenue generated by different products by year. Again we can see the shift of the trend in the sales of products as the solution is implemented in the current year.

- The *Total Revenue by Year* bar plot gives insights into the total sales revenue generated in each year. We can easily observe that the revenue generated has increased in the current year when the solution is implemented.
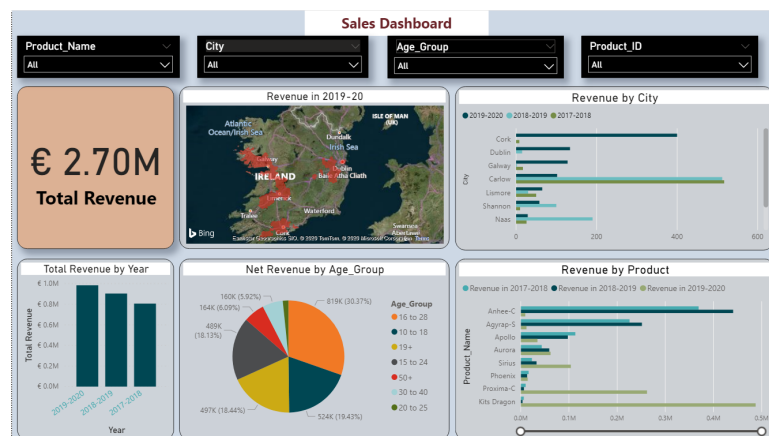
Figure 19: The sales dashboard

3.2.2 Customer Dasboard

The customer dashboard shows the distribution of no. of customers over the years by different products, cities and months. The selection switch on the top the dashboard enables user to view the data for any particular year, month and/or city (Refer to Figure 20).

- The *Monthly Customer Count* plot gives the no. of customers in each month in each of the three years. We can observe that the monthly customer count has increased in the current year substantially. This means that more potential customers(Leads) are now being converted to the consumers of the product.

- The *Customer Distribution By city* plot shows the no. of customers city wise over the years. As we have seen the sales dashboard, county Carlow has recorded the highest no. of customers for two consecutive years but the trend has changed it the current year.

- The *Total Number of Customers by Year* plot shows the net count of customers in each year. As we can see the count has increased in the current year after implementing the solution.

- The *City Wise Customer Distribution* map shows the distribution of customers in different cities in the current year.

Figure 20: The customers dashboard

3.2.3 Trends Dashboard

The trends dashboard basically highlights the trends or patterns in the sales of different products over the years (Refer to Figure 21).

- The *Total Revenue by Product* plot shows the net revenue generated by each product over all the three years combined.

- The *Revenue Trend by Product* plot provides insights into the sales of the individual product in each year.

- The three plots on the right displays the top three products that generated the highest revenue in each year.
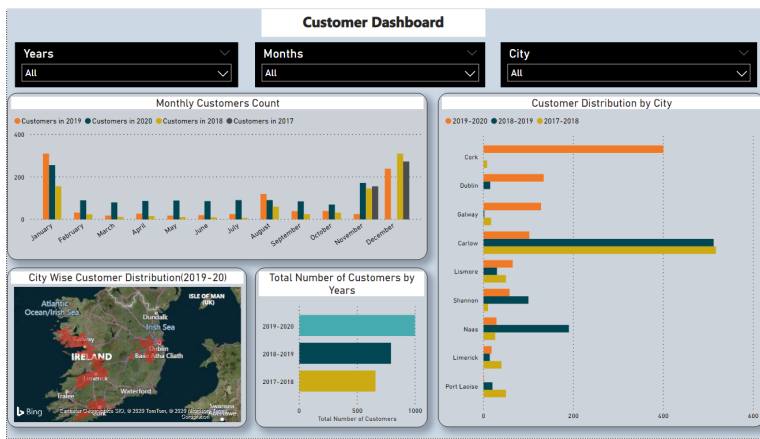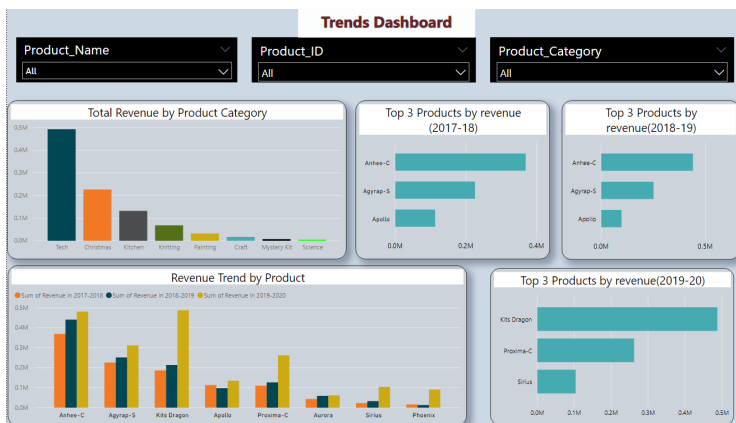

Figure 21: The trends dashboard

4 Benefits of the Solution

- As can be observed from the sales dashboard, the sales in the current year when the solution has been implemented has increased substantially as compared to the previous years. This means that more customers are actually engaging in the buying of DIY products. This also implies the higher number of customers being qualified from the Leads stage to the Invoice stage. Since the ongoing pandemic has forced people to sit in their homes, these products have proven to be an excellent productivity boosters for people. They not only learn but also build their own custom made products.

- Since the implementation of the solution, it can be seen that the target markets for the products have changed drastically. While in the previous two years, Carlow recorded the highest no. of sales and customers, inn the current year, the market has shifted to the major cities like Cork and Dublin which has eventually led to an increase in the sales revenue.

- The major advantage of the implementation of the solution is its robustness and efficiency. By migrating the customers data from a traditional database to a cloud based database, we have ensured that:

  – The data is fully encrypted and can only be accessed by the employees with a certain level of clearance.

  – The data is available from any location in the world, if needed.

  – The data is fail secure, meaning even if anything wrong happens with the company's server, the data will be safe in the cloud.

- By using the same model solution, the company's market can be expanded globally in major cities as well.

- Another major benefit of the solution is it's low cost and high flexibility. Since the cloud based services are cheap as compared to the traditional database management services, the company can use its resources in other domains where it is required that will aid the company in one of its objective to be a fully agile organization.

5 Conclusion

By observing various statistics and trends with the help of dashboards, the implemented solution can be considered to be a success. Although the revenue generated by some products show a negative trend, the overall revenue has increased after implementing the solution. This means that the company has to introduce some new products and make some changes in the current products over the coming years. The overall count of customers has increased substantially. The significant proportion of the company's has now been shifted to the major cities like Dublin.

6 Further Work

- The implemented solution make use of the Amazon's AWS cloud server, but equivalently any other cloud service like IBM, Azure can also be used for the same. It was used primarily because its subscription was provided free to the students. The company needs to decide carefully which cloud service to use considering the benefits and costs associated with each of them.

- The CRM software used for this project was Microsoft Dynamics 365, but equivalently any other CRM like sales force can also be used [2]. The main feature of Dynamics 365 used was the sales hub feature. But in future, features like Marketing, Finance, Service can also be explored to make the BI and BA system of the company even more flexible and robust.

- The company used for the purpose of this project was fictitious but it emulates most of the aspects of a real company. The Arecibo was created for just demonstrating the implementation of a working, scalable and lost cost BI solution that can be employed by an organization.

7 Appendix: Teamwork

- For the successful completion of the group project, we needed to make sure that a proper timeline was created, the workload would be equally divided, and that we had the access to a common communication medium. Microsoft Teams was used as the communication medium for the accomplishment of weekly goals that led to the successful completion of the project.

- Daily meetings were organised on the Microsoft Teams to discuss about the sprint plan.

- The whole project was divided into three major parts - Specification and the Implementation and the Presentation with each part containing various sub parts.

- Each part was divided into different sprints. 3 sprints were allotted for the specification part of the project and 5 sprints were allotted for the implementation part of the project. 2 sprints were allotted each for the writing of project report and preparing presentation. Each sprint had variable duration spanning from 2 days to 4 days.

- The successful and timely completion of the project was possible only because of each member of the group contributed sincerely their part into the project. All the team members assisted each other throughout the project.

References

[1] Robert S. Kaplan and David P. Norton. *The Balanced Scorecard*. Available: https://hbr.org/1992/01/the-balanced-scorecard-measures-that-drive-performance-2.

[2] *Microsoft Dynamics 365 Services*. Available: https://dynamics.microsoft.com/en-ie/.