**DECLARATION**

I hereby declare that the summer internship report (title) which is being submitted to **National Institute of Technology Delhi** in partial fulfilment of the requirements for the award of the degree of **B.Tech** in Computer Science has not been submitted to any University or Institution for the award of any degree to the best of my knowledge.

Name: Shruti Bhadoriya

Registration No.: 141100007

Computer Science Department

**MAQ Software**

MAQ Software was established on March 1, 2000 in Seattle, Washington to develop customized web solutions to meet the needs of enterprise customers.

Over the years, the company has been recognized by leading media in US and India for our ability to deliver great solutions.

MAQ Software specialises in Data Management and Reporting to help business users support their intuition with data. MAQ Software has worked closely with marketing, operations, and product groups across Fortune 500 companies. More than 50,000 Product, Sales, and Marketing Managers around the globe use custom applications (both on-premise and Cloud hosted), Line-of-Business (LOB), and Data Analytics & Business Intelligence (BI) solutions created and managed by organisation. Using the latest Agile engineering techniques in a disciplined manner, the company accelerates software initiatives that enable customers to transform their industries.

MAQ Software is a Preferred Vendor for the various prestigious Corporations, working closely with Marketing and Research and Development (R&D) teams in Redmond, Washington.

As a vendor, company works in collaboration with the ABC Pvt. Ltd. Corporation to support new marketing programs and technologies related to Windows and Office products. Our company has been acknowledged as one of the fastest growing companies in Washington State over the past several years.

As one of the fastest growing companies in Washington State, MAQ Software has expertise to ensure the success of customer’s software projects.

Quality, agile development, and cost savings drive clients to partner with MAQ Software. Clients can choose the work model that best meets their needs.

MAQ Software has accomplished the rare achievement of ranking on the Inc. 5000 list for the ninth time. Only 1% of Inc. 5000 companies demonstrate sustained growth by appearing nine or more times. The highly prestigious list has tracked the nation's fastest-growing private companies for 35 years.

Area of expertise:

* Data Management
* Artificial Intelligence
* Power BI
* App Development
* Cloud Transformation

**Chapter 1**

**Bootcamp**

2 months summer training commenced with an extensive boot camp on technologies which were relevant to the projects we would work on.

The major area of focus of the training was on :

* Transact-SQL
* Power BI

Following is the brief description about both the technologies.

**2.1 Transact-SQL**

Transact SQL or better known as T-SQL is a set of programming extensions from Sybase and Microsoft that add several features to the Structured Query Language ([SQL](http://searchsqlserver.techtarget.com/definition/SQL)), including transaction control, exception and error handling, row processing and declared variables.

T-SQL is used to interact with [relational databases](https://en.wikipedia.org/wiki/Relational_database). T-SQL expands on the SQL standard to include [procedural](https://en.wikipedia.org/wiki/Procedural_programming) programming, [local variables](https://en.wikipedia.org/wiki/Local_variable), various support functions for string processing, date processing, mathematics, etc. and changes to the [DELETE](https://en.wikipedia.org/wiki/Delete_(SQL)) and [UPDATE](https://en.wikipedia.org/wiki/Update_(SQL)) statements.

Transact-SQL is central to using [Microsoft SQL Server](https://en.wikipedia.org/wiki/Microsoft_SQL_Server). All applications that communicate with an instance of SQL Server do so by sending Transact-SQL statements to the server, regardless of the user interface of the application.

T-SQL's transaction and journaling system, handles just about anything - including a power cycle or hardware failure - without database corruption, and if something gets messed up it fixes it automatically.

T-SQL support CTE. A common table expression (CTE) can be thought of as a temporary result set that is defined within the execution scope of a single SELECT, INSERT, UPDATE, DELETE, or CREATE VIEW statement. A CTE is similar to a derived table in that it is not stored as an object and lasts only for the duration of the query. Unlike a derived table, a CTE can be self-referencing and can be referenced multiple times in the same query. It simplifies complex queries and most importantly enables you to recurse.

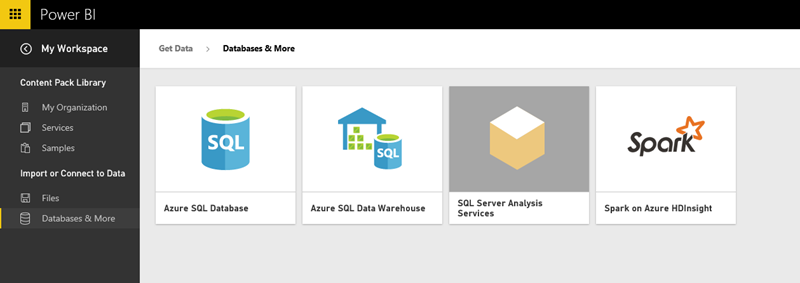
T-SQL is the SQL dialect that the product SQL Server is using. Transact-SQL is central to using SQL Server. All applications that communicate with an instance of SQL Server do so by sending Transact-SQL statements to the server, regardless of the user interface of the application. SQL Server is tied to Transact-SQL ([T-SQL](http://searchsqlserver.techtarget.com/definition/T-SQL)), an implementation of SQL from Microsoft that adds a set of proprietary programming extensions to the standard language.

**2.1.1** **Difference between T-SQL and SQL**

* T-SQL adds a number of features that are not available in SQL. This includes procedural programming elements and a local variable to provide more flexible control of how the application flows.
* A number of functions were also added to T-SQL to make it more powerful; functions for mathematical operations, string operations, date and time processing, and the like.
* These additions make T-SQL comply with the Turing completeness test, a test that determines the universality of a computing language. SQL is not Turing complete and is very limited in the scope of what it can do.
* Another significant difference between T-SQL and SQL is the changes done to the DELETE and UPDATE commands that are already available in SQL. With T-SQL, the DELETE and UPDATE commands both allow the inclusion of a FROM clause which allows the use of JOINs. This simplifies the filtering of records to easily pick out the entries that match a certain criteria unlike with SQL.
* SQL is non-procedural language since it deals with what data to be extracted. Whereas T-SQL is procedure language since it deals with what data to be executed and how it should be displayed.
* The SQL queries in SQL are submitted individually to the database server, while in T-SQL the batch program is written where in all commands are submitted to the server in a single go.

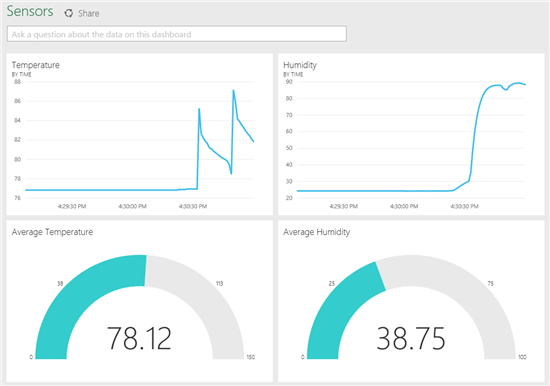
**2.2 Power BI**

Power BI is a cloud-based business analytics service from Microsoft that empowers anyone to experience any data – structured or unstructured – via simple drag-and-drop ease. Unlike many other dashboard solutions, Power BI can render live dashboards with moving charts and continuously updated visualizations for monitoring real-time streams from supported data sources.

****

**Figure 1: Power BI Connection**

As shown in figure 2, we can use the [Power BI REST API](https://powerbi.microsoft.com/en-us/developers) with any data source or Azure Streaming Analytics to render live Power BI Dashboards automatically. Alternatively, you can get near real-time analytics using simple “direct connect” data sources such as [Analysis Services](http://www.jenunderwood.com/2015/01/20/tip-how-to-power-bi-ssas/), Azure SQL Database, Azure SQL Data Warehouse or Spark with Power BI Reports.

****

**Figure 2: Power BI report**

As shown in fig 2, We can get any type of visuals from Power BI ,Microsoft Power BI tools makes the reporting part so easy that we can create our reports in short period of time with so much attractive and correct and real time data, milliseconds refresh data we can retrieve with our job scheduling.

You can import the data which is required to make the power BI report from various sources. It can be an excel sheet, CSV (comma separated value) file, database on your local machine, data in cloud etc.

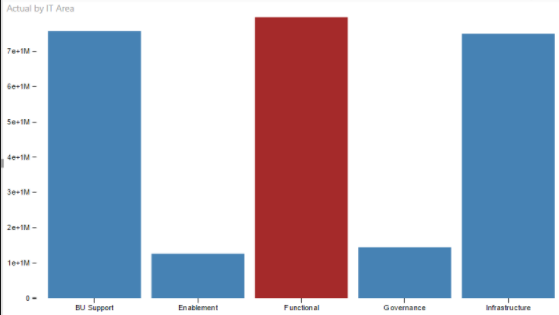
Power BI has Q&A feature to explore your data using intuitive, natural language capabilities and receive answers in the form of charts and graphs. Q&A is different from a search engine -- Q&A only provides results about the data in Power BI.

Data visualizations (aka visuals) helps us to interact with data to find business insights.

**2.2.1 Power BI visualizations**

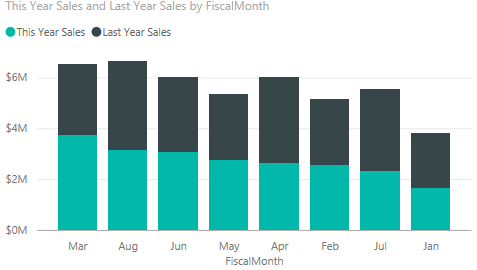
The types of visualization generally used in Power BI reports are:

1. Bar Chart



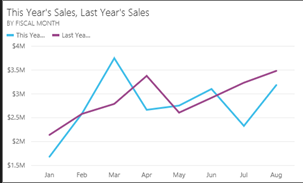
**Figure 3: Bar chart**

1. Stacked Bar Chart



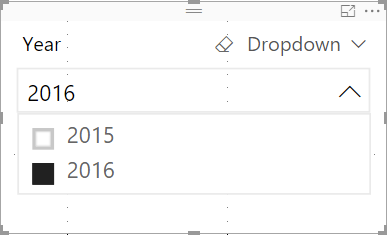
**Figure 4: Stacked Bar Chart**

1. Line Chart



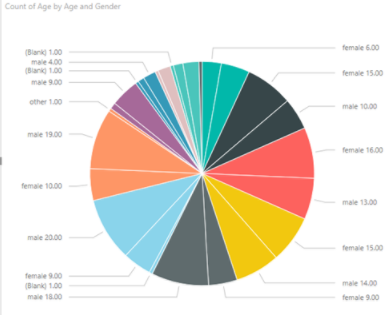
**Figure 5: Line Chart**

1. Slicer



**Figure 6: Slicer**

1. Pie Chart



**Figure 7: Pie Chart**

**Chapter 3**

**Technologies and Tools Used**

The technologies on which most of the work was Java, T-SQL, Power BI.

The tools which were used were Selenium Web Driver , SSIS, SSMS.

T-SQL and Power BI has been discussed before as a part of boot camp training, rest of the technologies and tools have been described below.

**3.1 Java**

**Java** is a general-purpose [computer programming language](https://en.wikipedia.org/wiki/Programming_language) that is [concurrent](https://en.wikipedia.org/wiki/Concurrent_computing), [class-based](https://en.wikipedia.org/wiki/Class-based_programming), [object-oriented](https://en.wikipedia.org/wiki/Object-oriented_programming), and specifically designed to have as few implementation dependencies as possible. It is intended to let application developers "[write once, run anywhere](https://en.wikipedia.org/wiki/Write_once,_run_anywhere)" (WORA), meaning that [compiled](https://en.wikipedia.org/wiki/Compiler) Java code can run on all platforms that support Java without the need for recompilation.

**3.2 SSIS (SQL Server Integration Services)**

**SQL Server Integration Services** (**SSIS**) is a component of the [Microsoft SQL Server](https://en.wikipedia.org/wiki/Microsoft_SQL_Server) database software that can be used to perform a broad range of [data migration](https://en.wikipedia.org/wiki/Data_migration) tasks.

SSIS is a platform for [data integration](https://en.wikipedia.org/wiki/Data_integration) and [workflow applications](https://en.wikipedia.org/wiki/Workflow_application). It features a [data warehousing](https://en.wikipedia.org/wiki/Data_warehouse) tool used for data [extraction, transformation, and loading (ETL)](https://en.wikipedia.org/wiki/Extract,_transform,_load). The tool may also be used to automate maintenance of SQL Server databases and updates to multidimensional [cube data](https://en.wikipedia.org/wiki/OLAP_cube).

**3.3 SSMS(SQL Server Management Studio)**

**SQL Server Management Studio** (SSMS) is a software application first launched with [Microsoft](https://en.wikipedia.org/wiki/Microsoft) [SQL Server 2005](https://en.wikipedia.org/wiki/Microsoft_SQL_Server) that is used for configuring, managing, and administering all components within [Microsoft SQL Server](https://en.wikipedia.org/wiki/Microsoft_SQL_Server). The tool includes both script editors and graphical tools which work with objects and features of the server.

**3.4 Selenium Web Driver**

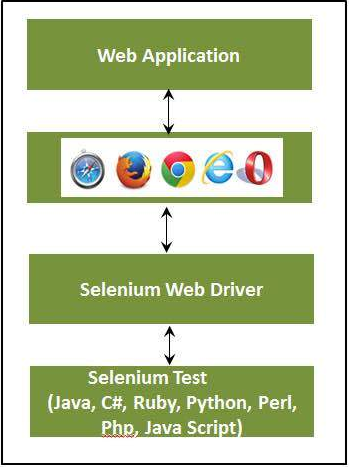
Selenium web driver was significantly used during the project, we mostly worked on the part of the project where the data pull process was to be automated.

**Selenium** is a portable [software-testing](https://en.wikipedia.org/wiki/Software_testing) [framework](https://en.wikipedia.org/wiki/Software_framework) for [web applications](https://en.wikipedia.org/wiki/Web_application).Selenium automatesbrowsers. Primarily, it is for automating web applications for testing purposes, but is certainly not limited to just that. Web-based administration tasks can be automated as well.

Selenium web driver is used in the scenario where:

* You want to create robust, browser-based regression automation suites and tests.
* You want to scale and distribute scripts across many environments.

[Selenium WebDriver](http://www.seleniumhq.org/projects/webdriver/) is a collection of language specific bindings to drive browser -- the way it is meant to be driven. Selenium WebDriver is the successor of [Selenium Remote Control](http://www.seleniumhq.org/projects/remote-control/) which has been officially deprecated.



**Figure 10: Web Driver Architecture**

**3.4.1 Selenium Locators**

Locator is a command that tells Selenium Web Driver which GUI elements (say Text Box, Buttons, Check Boxes etc) its needs to operate on. Different types of locators have been mentioned below.

**3.4.1.1 Locate by ID**

This is the most common way of locating elements since ID's are supposed to be unique for each element.

**Target Format:**id=id of the element

**3.4.1.2 Locate by Name**

Locating elements by name are very similar to locating by ID, except that we use the **"name="**prefix instead.

**Target Format:**name=name of the element

**3.4.1.3 Locate by Link Text**

This type of locator applies only to hyperlink texts. We access the link by prefixing our target with "link=" and then followed by the hyperlink text.

**Target Format**: link=link\_text

**3.4.1.4 Locate by CSS Selector**

**CSS Selectors are string patterns used to identify an element based on a combination of HTML tag, id, class, and attributes.**

CSS Selectors have many formats, but we will only focus on the most common ones.

* Tag and ID
* Tag and class
* Tag and attribute
* Tag, class, and attribute
* Inner text

**3.4.1.5 Locate by DOM (Document Object Model)**

The Document Object Model (DOM), in simple terms, is the way by which HTML elements are structured. Selenium can use the DOM in accessing page elements.

There are four basic ways to locate an element through DOM:

* getElementById
* getElementsByName
* dom:name (applies only to elements within a named form)
* dom:index

**3.4.1.6 Locate by Xpath**

XPath is the language used when locating XML (Extensible Markup Language) nodes. Since HTML can be thought of as an implementation of XML, we can also use[XPath](https://www.guru99.com/xpath-selenium.html)in locating HTML elements. It can access almost any element, even those without class, name, or id attributes.

There are two types of xpath

* Absolute XPath
* Relative Xpath

**3.4.2 Implicit wait and Explicit wait in Selenium**

**3.4.2.1 why wait is required in Selenium?**

Most of the web applications are developed using Ajax and Javascript. When a page is loaded by the browser the elements which we want to interact with may load at different time intervals.

Not only it makes this difficult to identify the element but also if the element is not located it will throw an "**ElementNotVisibleException**" exception. Using Waits, we can resolve this problem.

**3.4.2.2 Implicit Wait**

The implicit wait will tell to the web driver to wait for certain amount of time before it throws a "No Such Element Exception". The default setting is 0. Once we set the time, web driver will wait for that time before throwing an exception. implicit waits will be in place for the entire time the browser is open. This means that any search for elements on the page could take the time the implicit wait is set for.

**3.4.2.3 Explicit Wait**

An explicit wait is code you define to wait for a certain condition to occur before proceeding further in the code. The worst case of this is Thread.sleep(), which sets the condition to an exact time period to wait. There are some convenience methods provided that help you write code that will wait only as long as required. WebDriverWait in combination with ExpectedCondition is one way this can be accomplished.

Few of the expected condition are listed below:

* elementToBeClickable(By locator) : waits until an element is visible and enabled
* elementToBeSelected(WebElement element): waits until an element is selected
* presenceOfElementLocated(By locator) : waits until presence of an element
* textToBePresentInElement(By locator, String text): waits until specific text is present in the an element.

**List of Abbreviation**

* **LOB:** Line of Business
* **BI:** Business Intelligence
* **SDLC:** Software Development Life Cycle
* **SQL:** Structured Query Language
* **T-SQL :** Transact Structured Query Language
* **CTE:** Common Table Expression
* **CSV :**Comma Separated Values
* **API:** Application Programming Interface
* **SSIS :** SQL Server Integration Services
* **SSMS:**  SQL Server Management Sudio
* **WORA:** Write once, run anywhere
* **ETL:** Extract Transform and Load
* **DOM:** Document Object Model
* **HTML:** HyperText Markup Language
* **CSS:** Cascading Style Sheet
* **UI:** User Interface
* **JAR:** Java ARchive

**List of Figures**

* **Figure 1:** Power BI Connection
* **Figure 2:** Power BI report
* **Figure 3:** Bar Chart
* **Figure 4:** Stacked Bar Chart
* **Figure 5:** Line Chart
* **Figure 6:** Slicer
* **Figure 7:** Pie Chart
* **Figure 8:** Web Driver Architecture