Assignment 1 – ADSEI Project Diary

Team Members

- ANISH RAJ PANDEY - 217647455
- USMAN TARIQ SYED - 218062475
- PRATIK BHANDARI - 217416941
- ROHAIL AHMED KHAN - 217660287
- FAIQ Ul HASSAN - 218154944
- VENKATA LAKSHMI DURGA SEERAPU - 217499488
- MOHITH CHOWDARY PENDYALA - 218174968

Project Goal

To create a product, capable of allowing Students and Teachers to explore insights derived from visualizations of datasets in an interactive and engaging way.

Project Value

Value Hypothesis

The product aims to provide High School Students, Teachers and other non-technical individuals with an interactive learning environment where they are presented with dynamic, Visual representations of Datasets and their corresponding derived insights. It aims to teach the value of Data Science being applied in real world situations.

Growth Hypothesis

The project has a vast scope when it comes to portraying the teachings of Data Science. It can incorporate the current methods which are being used in the industry by professionals to derive advanced insights from datasets which usually are overlooked by others. A simple example of this would be to predict the market price of an item in the coming couple of weeks and many others.

Metrics

Initially, we will assess the success of our project by being assessed for the initial project deliverables by the client (Dr Linda McIver) and our assigned supervisor (Dr Gang Li). For the later part of the project, we will maintain a Metrics document in which we will strictly keep track of project requirements being met by the project deliverables.

Project Plan

Communication Strategy

We will incorporate the following communication tools to organize effective internal communication for the project.

1. Emails
   - Primary form of communication with the client and supervisor to put forth queries, ask for feedback and discuss project progress and deliverables.
2. Meetings
   - Weekly-in person meeting of the team with the supervisor to assess project status and set project expectations.
   - Fortnightly video conference meeting with the client to discuss project deliverables and keep update of the project progress.
3. Team Collaboration Software
   - Trello: To assign tasks for members, set up deadlines, upload project files and links for sharing and record team activity as well as task completion status of all team members.
   - Slack: To share information and tools and to be in constant communication with team members and supervisor.
4. Github
   - Use access control and task management tools to maintain the project in the repository. Also, use links for sharing and create lists for issues to maintain the project by all members.

Task Management Strategy

The team comprises of 7 members where the basic workload can be summed up as –

1. Team Lead – Responsible for maintaining smooth information flow, facilitating the team, obtaining the required resources and resolving team conflict and problems.

2. Product Owner - Responsible for gathering requirements, developing an overall vision of the product and conveying that vision to the team.

3. UX/UI designer – Responsible for research, wireframe creation, task analysis, creating prototypes and UX testing.

4. Front-end developers – Responsible for implementing visual components which the users see and interact with in the web project by implementing bootstrapping strategies and D3 visualization tool.

5. Back-end developers - Responsible for database creation, integration and management of the entire back-end function and primarily involved in adding utility to everything that front-end developers create.

We will keep track of the project's lifecycle by focusing on different stages of the product development cycle:

1. **Project Proposal**: We're going to assess what the basic requirements of the products are and what it aims to achieve from a business point of view. We will form up a proposal which includes Project Scope document, Requirement gathering techniques and receiving approval from our supervisor and client.

2. **Project Planning & Design:** In the planning stage we're going to identify and divide the tasks amongst team members. We're going to design initial designs for the web application and it functionalities. We will select relevant datasets and technologies which will help us in delivering the goals of the product.

3. **Project Execution:** In this stage we're going to start the development procedure of the product which includes the development of the Frontend, Backend, Visualization tool and integration of the dataset with the actual Web application. All deliverables will be carefully checked with a Quality Metrics document to ensure that all requirements are being met and there are is minimum compromise on any quality related issues.

4. **Project Closeout**: This stage is related to the closure of the project which includes thorough product testing, project deployment and project handover to the client.

Task Management Strategy

Describe here how do you divide the work among the project team members and how will you keep track of the project lifecycle.

Iteration Plans

Iteration 1	Product Initial Start		
Start Date	26/07/2018	End Date	05/08/2018
Description	First introductory meeting with client. Strategize attack plan for project start		
Result	Basic Idea of expected Product and formation of Project Strategy.		
Follow Up	Start to formulate queries for requirements and research available resources.		

Iteration 2	Project Planning Stage		
Start Date	06/08/2018	Start Date	12/08/2018
Description	Selection of dataset along with development of a simple website for the product.		
Result	Concludes initial project planning stage		
Next Step	Develop initial prototype for demonstration and feedback		

Iteration 3	Prototype Planning & Development		
Start Date	13/08/2018	End Date	25/08/2018
Description	Plan and develop first prototype of project, including website & visualization tool.		
Result	Initial Prototype ready for assessment		
Follow Up	Work on improving prototype according to received feedback		

| Iteration 4	Development Stage		
Start Date	27/08/2018	End Date	09/09/2018
Description	Implementing dataset into application and visualization tool. Integrating Backend functionalities into the application		
Result	Achieving project deliverables.		
Follow Up	Plan for Testing and Deployment stage		

Iteration 5	Test and Deployment Stage of Project		
Start Date	10/09/2018	End Date	18/09/2018
Description	Testing all functionalities and Project Deliverables. Making sure everything meets Requirement Specifications. Deploying the Product on client's servers		
Result	Deliverables meet the Specified Requirements. Deploying the Product.		
Next Step	Project Closure		

Design

The main idea behind the product is to utilize any type of open source datasets which are freely available to us on the internet and apply Visualization techniques which support any facts corresponding to the given dataset. We're also required to create a dynamic interactive tool for users to utilize which allows them to understand the data through dynamic graphical representations.

Architecture

Overview
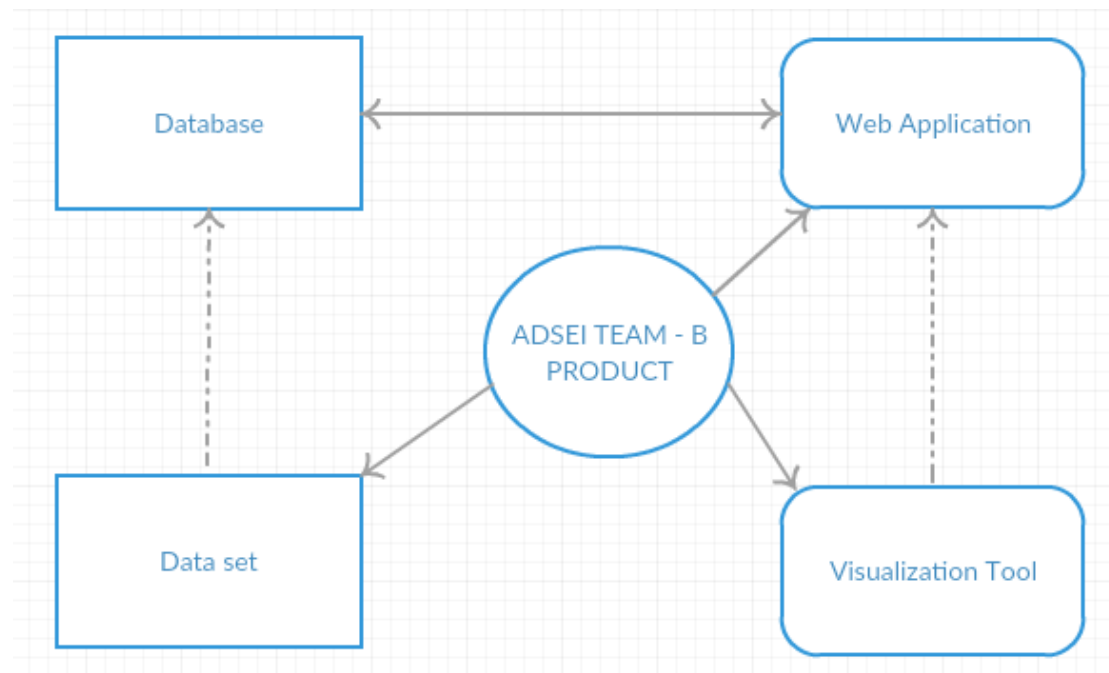
The ADSEI product that we plan to complete (at the very initial stage) consists of 4 different parts, which can be seen from the diagram above. Firstly, the product will consist of a Web Application where users will interact with the website on which the actual visualization tool is displayed on.

Secondly, we will design the Visualization Tool which will be integrated into the Web Application so that the Users can observe meaningful insights on the corresponding dataset. This tool will be made dynamic allowing the users to view different types of graphical representations in different display formats.

We will integrate our selected dataset with the database technology that we'll be using which in turn will be implemented into the actual web application to be utilized by the Visualization Tool.

Architectural Decisions

1	Dataset Selection
Problem Statement	Our product requires us to select a suitable online open sourced dataset on which we can perform factual Visualization techniques
Available Options	
Option 1.	**Climate Change: Earth Surface Temperature Data** Description: This dataset provides us with a large repository of the earth's Temperature which are further categorized into Countries, States and Cities. The Dataset also

	consists of attributes which provide us with min & max temperature and many more. It also contains records which are spanning over 1900s till 2010s
	Pros:
	- Contains categorized data. - Can be used to display data over a time series pattern. - Contains continuous data which could be used for predicting future temperature as an added functionality. - Open Sourced.
	Cons:
	Dataset is very large.
Option 2.	**Weather Dataset**
	Description:
	A weather dataset which has data records containing observations from multiple Australian weather stations
	Pros:
	- Can be used to display data over a time series pattern Ease of availability of product implementing this model. - Contains continuous data which could be used for predicting rainfall in future days. - Open Sourced
	Cons:
	Data is not categorized and will require a lot of cleaning.
Option 3.	**Australia NSW Traffic dataset**
	Description:
	Provides data regarding different types of recorded Traffic fines in the NSW state.
	Pros:
	- Contains both Continuous and Categorized data which can be used to derive visual facts and be used for predicting purposes. - Open Sourced
	Cons:
	- Requires lots of effort in cleaning and organizing the dataset. - Extremely large.
Decision	
Selected Option	The selected option is **Option 1.**
Justification	Option 1 is the best case in our scenario as it allows us to work without the need to put in extra effort in cleaning the data and has an excellent scope which is perfectly tailored for the project.
Implications	We will need to apply cleaning techniques in the dataset for missing or misleading records.

2	Visualization Tools
Problem Statement	Our Client requires us to utilize Visualization tools which can provide dynamic viewing functionality
Available Options	
Option 1.	**D3.JS** Description: D3.js is a JavaScript library for producing dynamic, interactive data visualizations in web browsers. It makes use of the widely implemented SVG, HTML5, and CSS standards. Pros: • Documentation is easily available and easy to comprehend. • Can be used to display data over multiple visualized formats. • Has JavaScript Syntax which is easy to learn and apply. • Has many examples which are available online. • Open Sourced. Cons: • Will require some time to learn the tool.
Option 2.	**Python's Seaborn** Description: Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. Pros: • Documentation is easily available. • Can be used to display data over multiple visualized formats. • Has may example which are available online. • Open Sourced Cons: • Will need to be familiar with Python language to use. • Is a bit complicated to learn.
Option 3.	**Python's Geoplotlib** Description: Geoplotlib is a python toolbox for visualizing geographical data and making maps. Pros: • Powerful when trying to make maps. • Can be used to display data over multiple visualized formats. • Open Sourced Cons: • Will need to be familiar with Python language to use. • Is a bit complicated to learn. Few examples are available online.
Decision	

Selected Option	The selected option is **Option 1.**
Justification	Option 1 is the best case in our scenario as it has lots of documentation available online and is easy to learn compared to other options. It also has functionalities which are more suitable for our project.
Implications	We will need to start going through the documentation and tutorials to learn it as soon as possible.

3	Database Technology Selection
Problem Statement	Our product requires us to utilize the services of a database to store our dataset
Available Options	
Option 1.	**IBM CLOUD** Description: IBM Cloud is a cloud computing service which provides data storage services. Pros: - Many team members are familiar with it. - Documentation is easily available. - Suitable to store large data repositories. - Has many options for storing and organizing data repositories Cons: - Might require some time to familiarize with.
Option 2.	**MySQL Server** Description: A Relational Database Management System which is open sourced. Pros: - Some team members are familiar with it. - Documentation is easily available. - Low learning curve. - Open Sourced Cons: - Might not be suitable to store large datasets. - Not flexible when it comes to options for storing and organizing data repositories.
Decision	
Selected Option	The selected option is **Option 1.**
Justification	Option 1 is the best case in our scenario as many team members are familiar with its services which are more suitable for our project.
Implications	We will need to start learning the different required services for storing data.

Retrospective

Up till now we have been able to have a clear picture of what the project is about and what type of a product the client is expecting from us. After having the first meeting, we strategized amongst ourselves to devise an attack plan for the coming project. We established clear communication amongst the team members and effectively communicated on what is to be achieved and more importantly, How? We planned to start the project initially by researching different online datasets that were available to us and then we selected the most suitable one out of them. We decided on the different technologies that we were going to need to successfully complete the project. And lastly, we have identified and broken down the coming tasks by categorizing them into the different phases of a Software Development Life Cycle (SDLC) and prioritized them correspondingly into multiple iterations in chronological order.