

## Introduction:

Student Information System is a software which is used to manage the student-related data. Different data of student like admission details, pending bills, attendance records, grade reports, performance history, and coursework planning are recorded in the Student Information System. The data related to teachers like student graduation rate, assigned classes and future coursework planning along with course materials are also recorded in the Student Information System. Student Information System is known as various names like School Management Software, Student Administration System or even SIS as it's an acronym. No matter how many different names it has, it's primary work is the management of student information. The student data is managed in a unified form so that it will be easy to access the data.

The process of collecting and organizing data into a common database is known as Data Warehousing. It will help in supporting business goals and organize end-user information. "Data Warehousing incorporates data stores and conceptual, logical, and physical models to support business goals." (Bose, 2018) A data warehouse is created to provide a long-term interpretation of data over time. Data warehouse offers a single, complete source of present and past information. (Bose, 2018) This can be used in the student information system to perform analysis in the student and teacher-related data from various sources. It also helps to manage the storage by the help of ETL (Extraction, Transformation, and Loading) process.

Data visualization is used to create a simple graphical representation from complex data. Visualization is vaguely used in data analysis, especially for data forecasting and predictions. Visualization plays a vital role in analysis and reporting as our brain processes image 60,000 times quicker than texts (Eisenberg, 2014). So instead of pages and pages of reports, a simple chart that can convey all of the information in an instant is preferred. In a student information system, instead of going through heaps of data of a student, a simple chart showing the progress or attendance of a student can be created so that the declining students can be determined and the actions can be taken immediately.



Figure 1: Data warehousing for analysis (DataFloq, n.d.)

#### Research Questions:

- Does data visualization makes analyzing a student's progress easier?
- Is implementation of data warehousing techniques in SIS instead of traditional spreadsheet-based SIS at Softwarica College a right choice?
- Will implementing data warehousing and visualization techniques improve the graduation rate at Softwarica College?

#### Aim of the project:

**Implement different data warehousing and visualization techniques in Student Information System to help educational institute improve their student graduation rate**

#### Objectives of the project:

##### Personal Objective:

- Implement the knowledge gained through various research papers and classes.
- Design a proper database
- Add quality features in the project
- Show eagerness, creativity and commitment of the task

##### Academic Objective:

- Meet the deadline of the project
- Complete the research
- Prepare a standard document

##### Technical Objective:

- Create a Linux environment for Oracle 11g
- Prepare a reliable, secure and robust database
- Enable data recovery and backup features

- Produce easy-to-understand charts
- Implement data warehousing technologies

### Justification of the project:

This existing system doesn't show the progress of the students. The data isn't utilized properly. This existing system can be further enhanced by using different data warehousing techniques and visualization techniques. With the help of different data warehousing techniques, the student data can be managed more efficiently, and analysis can be done to predict the future results of the student. The different visualization techniques can be used to graphically represent the result obtained from the analysis of data. The visualization techniques help to study the result more easily and accordingly make a decision.

So, the enhanced version of SIS will help in making a predictive model and acts as a decision support system. This system will help the educational institute to decide on which students need attention and make plans to support them.

Consider a scenario shown in the figure below. The employee is tired of managing data from spreadsheet software. He feels stupid and is wondering if there is any better way to do this work.



Figure 2: Problem

Then as shown in figure below, the business owners are having a meeting where they are putting out their problems to the DBA and the DBA is recommending the possible solutions that can be implemented as the solution for the owners problems.



Figure 3: Discussion for appropriate solution

In the figure below, the features of the oracle are shown which helped to determine oracle as the most appropriate solution in this scenario.



Figure 4: Solution

Finally, as shown in the diagram below, the staff is happy as it is now easier to manage and report the data to others.



*Figure 5: Solution implemented*

### Keywords:

Oracle; Database; Data warehouse; Data Visualization; Student Information System (SIS); Tableau;

### Project Title:

Application of multiple Data warehousing and Visualization techniques in Student Information System

### Client, Audience and Motivation:

Student Information System is used actively by almost all colleges and schools to manage the student-related data. Softwarica College of IT & E-Commerce is a college located at Dillibazar, Kathmandu, Nepal. This college also uses a student information system to manage student-related data like admission details, pending bills, student results. But the student information system lacks visualization features and data warehousing is also not used in the system. So, I have selected Softwarica College of IT & E-Commerce as the client for this project.

This project tends to manage the data properly and use various visualization techniques to produce charts from the available data. By doing so, the concerned party (administrative and academic staffs) can predict the outcome of a student and provide attention to weaker students to improve the student graduation rate. Also, the coordinators can monitor the performance of a teacher and decide the most critical subjects to provide more attention to those subjects. This will ultimately contribute to the betterment of the student's education and management team's easiness and to improve the response time.

### Mini Literature Review:

In a study conducted by Vivienne V. Forrester, he claims that even though the School Management Information Systems uses databases, they are **simply Information Storage Systems**. They are incapable in making the relevant data readily accessible. Despite the usage of several databases in schools for various activities like attendance, grades, test scores and payment records, the databases can hardly ever interconnect. This is leading to failure in making the appropriate use of data to make decisions intended for the students. So, he has proposed a conceptual model that will help in solving the persistent issue related to data.

His study provides just a conceptual model, and theory on the steps for the proposed conceptual model. He has not mentioned about the visualization and its impact on decision making. Also, his model has not been implemented so the success or failure of the model is undetermined.

I will be attempting to implement visualization and different data warehousing techniques in the Student Information System to determine the success or failure of its implementation and to determine the impact of the visualization on decision making of both academic and administrative staffs.



### Primary Research Plan:

Data warehousing and visualization will be used in the Student Information System of Softwarica College. Then the result will be recorded and analysis will be performed on it. Also, an interview and questionnaire to the concerned administrative and academic staffs will be conducted to know their opinion towards the difference in the Student Information System and their preference.

This will be done through following steps:

- Deploy different data warehousing and visualization techniques in Student Information System at Softwarica College
- Collect the opinion and preference of the administrative and academic staffs
- Compare the student graduation rate before and after using data warehouse and visualization
- Summarize and report the final findings of the research

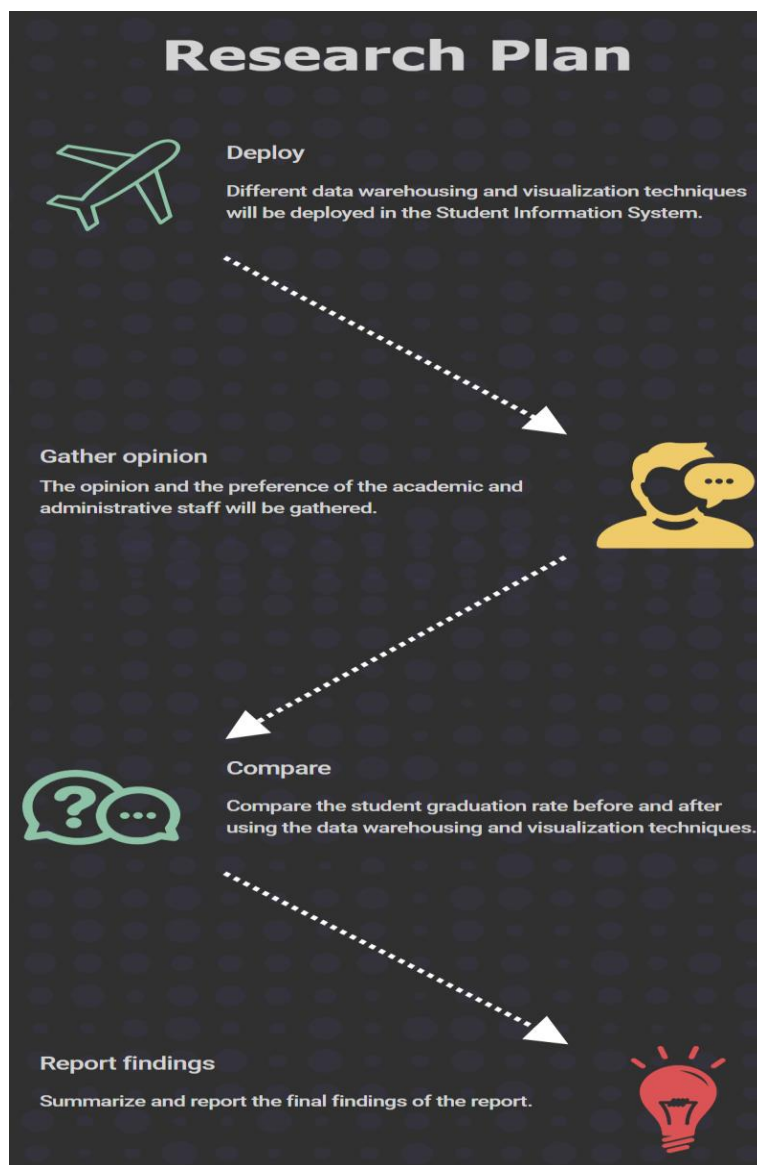


Figure 6: Research plan

## Methodology:

Methodology are sets of methods used to perform an activity or a study. ([merriam-webster, n.d.](#)) They make the tasks more organized and structured. There are different methodologies for different tasks. The most popular and appropriate methodologies for project development are:

- CRISP-DM
- Agile
- DSDM
- Prototyping

Among the above listed methodologies, I have selected CRISP-DM as it most appropriate for a database related project. CRISP-DM is the acronym for Cross Industry Standard Process for Data Mining. It provides an organized process for developing a project. It consists of six phases which can go into reverse so that the tasks in previous phases can be repeated. The six phases of CRISP-DM along with their sub-tasks are as follows:

1. Business understanding
  - Determine project outcomes: Objectives and success criteria for the project
  - Evaluate current condition: Requirement analysis, resources required for the project
  - Determine data mining goals: Determine the criteria to measure success of data mining
  - Create a project plan: Stages of the project along with their timeline and dependencies
2. Data understanding
  - Data collection: Gain the data required for the project
  - Data Description: Assess if the data satisfies the requirement
  - Data exploration: Find relationships between attributes, perform simple analysis
  - Data quality verification: Assess the completeness and correctness of the data
3. Data preparation
  - Select data: Decide which data is going to be used for the project
  - Clean the selected data: Remove unwanted attributes from the selected data
  - Create essential data: Derive values from existing data to create new attributes
  - Integrate data: Perform aggregations in the data or merge data
4. Modeling
  - Select the modelling technique: Decide the modelling technique to be used in the project
  - Generate test design: Test the quality and correctness of the model
  - Build model: Create the model from the selected datasets
  - Evaluate model: Evaluate the model as per the determined success criteria and rank them
5. Evaluation
  - Evaluate the results: Evaluate the models on the basis of business objectives
  - Review process: Find the tasks that were missed out and that needs to be repeated.
  - Decide on next steps: Determine the possible actions that will be done

## 6. Deployment

- Deployment plan: Determine how the deployment will be done and the steps that will be performed
- Monitor and maintenance plan: Create a plan for monitoring and maintenance
- Create a final report: Final report containing the summary, experience and result of the project
- Review the project: Evaluate the success and failure of project and places for improvements

(SmartVision, n.d.)

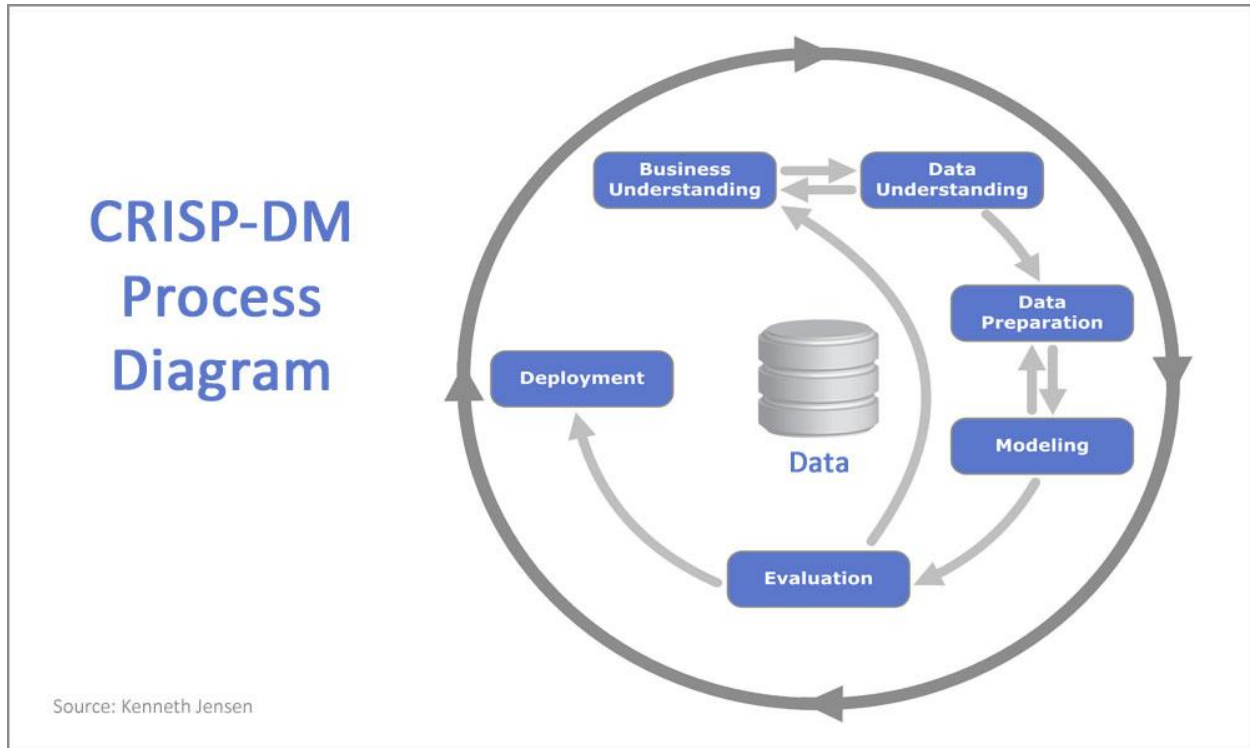










Figure 7: CRISP-DM process diagram

### Tools:

Various tools will be used for the completion of this research. Some of the tools that are essential are as follows:

Software	Icon	Version
Oracle VM VirtualBox		5.2.4
Putty		0.71
SQL Tools		1.6 Beta build 15
Microsoft Word		2016
Microsoft Excel		2016
Microsoft Project		2016
Photoshop		CS6
Notepad++		7.7.1

### Technologies:

Oracle was selected as the appropriate solution as there are multiple technologies within it which can be used to enhance the product. Some of the technologies that will be used are shown in the word cloud below:



Figure 8: Technologies used

## Deliverables:

Deliverables are simply the outcomes that will be received at the end of a specific period. Since the project is 4 months long (16 weeks), for the success of the project the following deliverables along with their deliverable weeks are shown in the chart below:

Week	Deliverable
01	<b>Project Title Research</b> A research will be done on deciding the project title. Title will be presented to supervisor for further improvement.
02	<b>Final Project Title</b> After constant feedback, final project title will be determined and submitted.
03 04	<b>Draft Proposal</b> A draft proposal for the project will be presented to the supervisor for feedback.
05	<b>Research Question</b> Research question essential for the report will be identified.
06	<b>Project Planning</b> A proper planning showing all the timelines will be prepared.
07	<b>Final Proposal</b> After constant feedback, a final proposal will be submitted.
08 09	<b>Product Preparation</b> The product with all the essentials features will be prepared.
10	<b>Visualization</b> Visualization will be performed on the data and various charts will be prepared.
11 12 13	<b>Research</b> Research will be conducted to find the answers of the research questions defined on previous weeks.
14 15	<b>Draft Research Paper</b> Draft research paper will be submitted for feedback and changes will be made accordingly.
16	<b>Final Research Paper</b> After constant feedback and improvements, final research paper will be submitted.

Figure 9: Deliverables

## Project Plan:

Project plan is the plan developed for the successful completion of a project within the determined timeframe and contains the stages of the project along with their timeline and the dependencies of the stages. For this project the following project plan is developed for the timely and successful completion of the project:


























	Task Mode ▾	Task Name ▾	Duration ▾	Start ▾	Finish ▾	Predecessors ▾
		<b>Proposal</b>	<b>23 days</b>	<b>Mon 7/22/19</b>	<b>Wed 8/21/19</b>	
		Start	1 day	Mon 7/22/19	Mon 7/22/19	
		Project Title Research	5 days	Tue 7/23/19	Mon 7/29/19	2
		Final Project Title with Aims and Objectives	3 days	Tue 7/30/19	Thu 8/1/19	3
		Draft Proposal	3 days	Fri 8/2/19	Tue 8/6/19	4
		Research Questions	2 days	Wed 8/7/19	Thu 8/8/19	5
		Project Planning	3 days	Fri 8/9/19	Tue 8/13/19	6
		Final Proposal	6 days	Wed 8/14/19	Wed 8/21/19	7
		<b>Development</b>	<b>32 days</b>	<b>Thu 8/22/19</b>	<b>Fri 10/4/19</b>	
		Oracle Installation	1 day	Thu 8/22/19	Thu 8/22/19	8
		Prepare Database	2 days	Fri 8/23/19	Mon 8/26/19	10
		Additional features	5 days	Tue 8/27/19	Mon 9/2/19	11
		Data Preparation	9 days	Tue 9/3/19	Fri 9/13/19	12
		Data Warehousing	5 days	Mon 9/16/19	Fri 9/20/19	13
		Visualization	10 days	Mon 9/23/19	Fri 10/4/19	14
		<b>Research</b>	<b>26 days</b>	<b>Mon 10/7/19</b>	<b>Mon 11/11/19</b>	
		Deploy	3 days	Mon 10/7/19	Wed 10/9/19	15
		Gather Opinions	3 days	Thu 10/10/19	Mon 10/14/19	17
		Analysis	12 days	Tue 10/15/19	Wed 10/30/19	18
		Compare Outcome	3 days	Thu 10/31/19	Mon 11/4/19	19
		Report Findings	5 days	Tue 11/5/19	Mon 11/11/19	20
		<b>Documentation</b>	<b>20 days</b>	<b>Tue 11/12/19</b>	<b>Mon 12/9/19</b>	
		Draft Research Paper	13 days	Tue 11/12/19	Thu 11/28/19	21
		Final Research Paper	7 days	Fri 11/29/19	Mon 12/9/19	23

Figure 10: Project Plan

The project plan shown in the figure above shows the tasks along with the sub-tasks that is required to be completed in the planned time frame for the completeness of the project. Also, the predecessor column shows which tasks need to be completed before the another. In the above project plan the working days are from Monday to Friday i.e. 5 days a week.

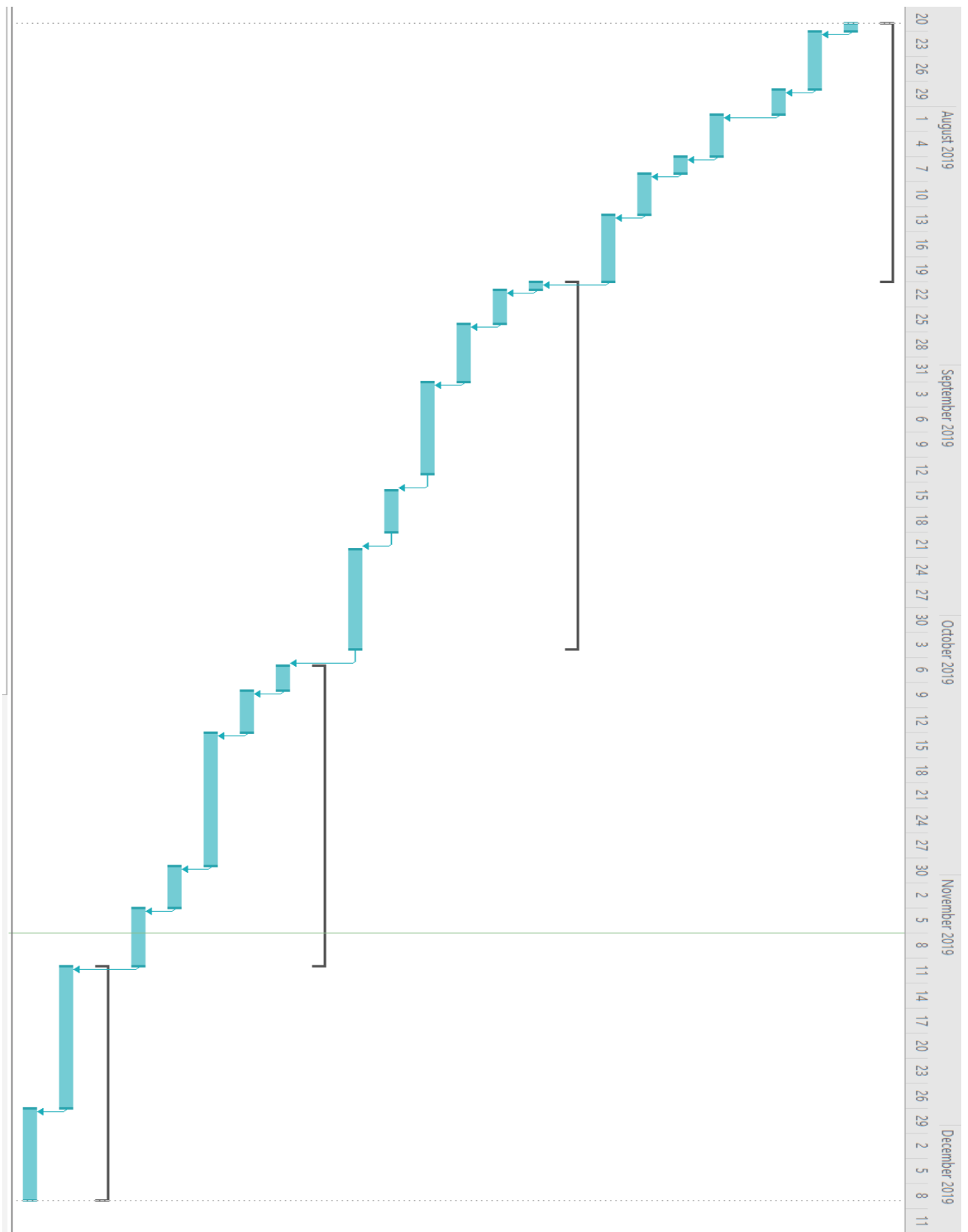


Figure 11: Gantt Chart