VIETNAM NATIONAL UNIVERSITY - HO CHI MINH CITY
INTERNATIONAL UNIVERSITY

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

DATA SCIENCE & DATA VISUALIZATION

PROJECT PROPOSAL

TOPIC QUESTION:

HOW AMBIENT FACTORS AFFECT STUDENT'S STUDYING OUTCOMES

BY GROUP 03

1. TRUONG DANG KHOA - ITDSIU19027
2. VU VIET PHONG – ITDSIU19048
3. TRAN NGUYEN KHANH DUY – ITDSIU18049
4. DUONG TRAN NHAT MINH – ITDSIU20032

TABLE OF CONTENTS

Github repository link: https://github.com/vtenpo/DSDV_Project

I. INTRODUCTION:

1. Abstract

The acknowledgment of factors affecting students' learning outcomes is undeniably essential to every educational institution. With this knowledge, schools, teachers, and parents can assist students in studying more efficiently and increasing their performance in courses. This project was created due to this reason - to answer the question "How ambient factors affect student's studying outcomes." The project's product is a website that provides interactive charts to help users visualize and find the answer to the above question. All the data used for the visualization purpose of this project is from Kaggle - an online community of data scientists and machine learning practitioners.

This project is established as a requirement for the Data Science and Visualization course at International University - Vietnam National University. The aim of the project is to practice data analyzing and visualizing skills and learn how to create a website with HTML, CSS, and JavaScript at the elementary level. The final product of this project should be a website with functioning interactive charts, and the data should be visualized in the most efficient way possible.

2. Purpose

By analyzing and visualizing this dataset, we could get a deeper insight into the influence of parents' education background, test preparation, ethnicity, and gender on students' performance. Moreover, we can come up with a better approach to help the students get a higher score; more importantly, it's the deeper understanding of the knowledge.

II. DATA PULLING PROCESS

1. About the sources

This data comes from Kaggle. What is Kaggle? Kaggle is an online community of data scientists and machine learning practitioners. Kaggle permits customers to locate and put up records sets, discover and construct fashions in a web-primarily based totally records-technological know-how environment, work with other data scientists and machine learning engineers, and enter competitions to solve data

science challenges. We decided to focus on data choose this dataset because it contains marks of students in math, reading and writing by the students in high school by Students from the United States. The main objective of analyzing this data is to understand which factors influence students' performance the most. The variables considered are race, the level of education of the parents, diet and the way in which the students prepared for the exams.

2. Data file

This dataset is composed of the following variables:

- Gender: Male or female
- Race/ethnicity: Grouped from A to E
- Parental level of education: Grouped from high school to master's degree
- Lunch: Type of lunch (standard or reduced)
- Test preparation course: If a student did the test preparation course before the exams
- Math score
- Reading and Writing score

	A	B	C	D	E	F	G	H
1	gender	race/ethnicity	parental level of education	lunch	test preparation	math score	reading score	writing score
2	female	group B	bachelor's degree	standard	none	72	72	74
3	female	group C	some college	standard	completed	69	90	88
4	female	group B	master's degree	standard	none	90	95	93
5	male	group A	associate's degree	free/reduced	none	47	57	44
6	male	group C	some college	standard	none	76	78	75
7	female	group B	associate's degree	standard	none	71	83	78
8	female	group B	some college	standard	completed	88	95	92
9	male	group B	some college	free/reduced	none	40	43	39
10	male	group D	high school	free/reduced	completed	64	64	67
11	female	group B	high school	free/reduced	none	38	60	50
12	male	group C	associate's degree	standard	none	58	54	52
13	male	group D	associate's degree	standard	none	40	52	43
14	female	group B	high school	standard	none	65	81	73
15	male	group A	some college	standard	completed	78	72	70
16	female	group A	master's degree	standard	none	50	53	58
17	female	group C	some high school	standard	none	69	75	78
18	male	group C	high school	standard	none	88	89	86
19	female	group B	some high school	free/reduced	none	18	32	28
20	male	group C	master's degree	free/reduced	completed	46	42	46
21	female	group C	associate's degree	free/reduced	none	54	58	61

III. PROCESSING AND VISUALIZING

1. Visualization design

This dataset includes scores guaranteed by students in various subjects. This analysis aims to find answers to understand the influence of important factors such as parents' education level, exam preparation status, etc. on students' results in exams based into two groups of subjects of logical thinking and linguistic thinking. From this data set, we divided into 2 subject groups. Group 1 is Math; Group 2 includes Reading and Writing. Each group will be divided according to the score level including Excellent, Good, Average, Below Average. Our job is to analyze the influence of the above factors on each score group of each subject.

2. Features

The website will consist of two main pie charts showing the distribution of scores. Each chart will have basic interactive features. Besides that, when the user clicks on the graph, specific details will be shown.

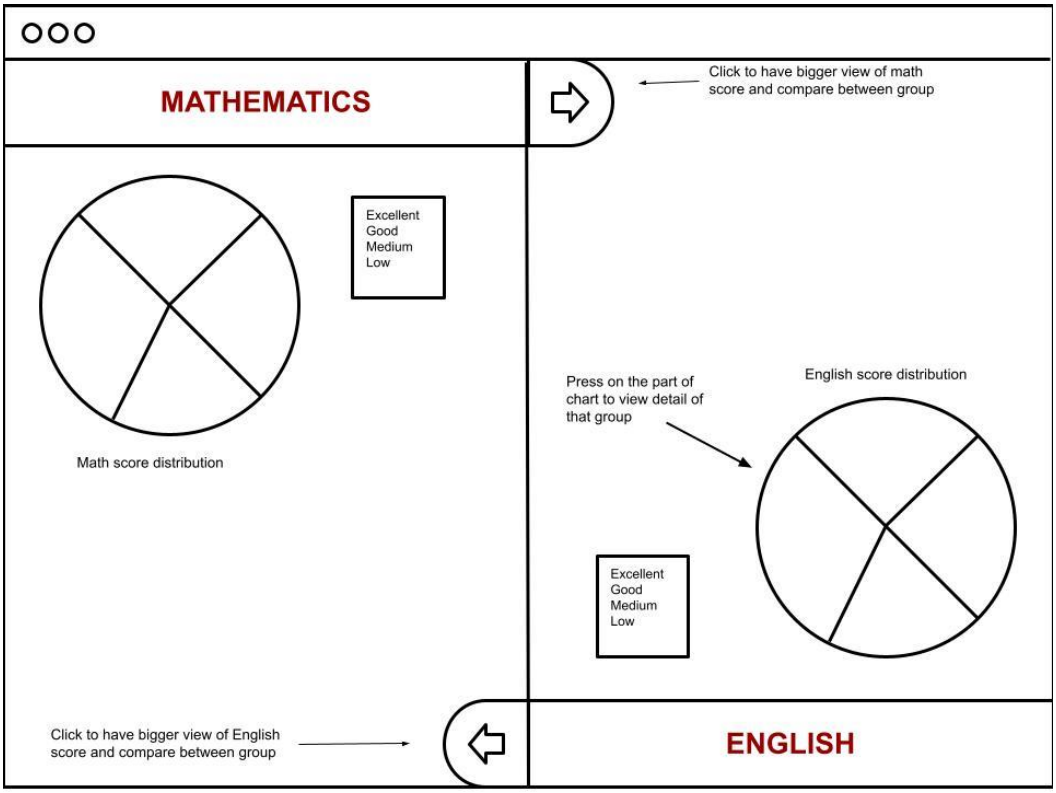
We are also thinking about the "comparing" feature so that the users can view and compare between groups.

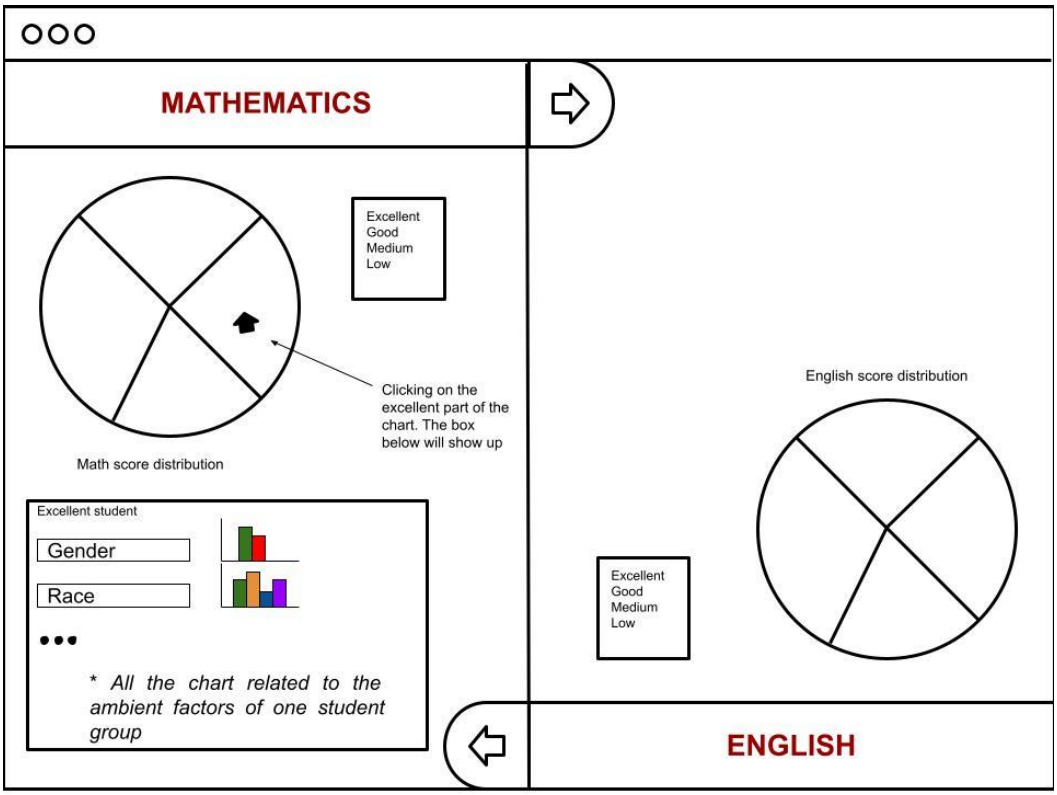
IV. PROJECT TIMELINE

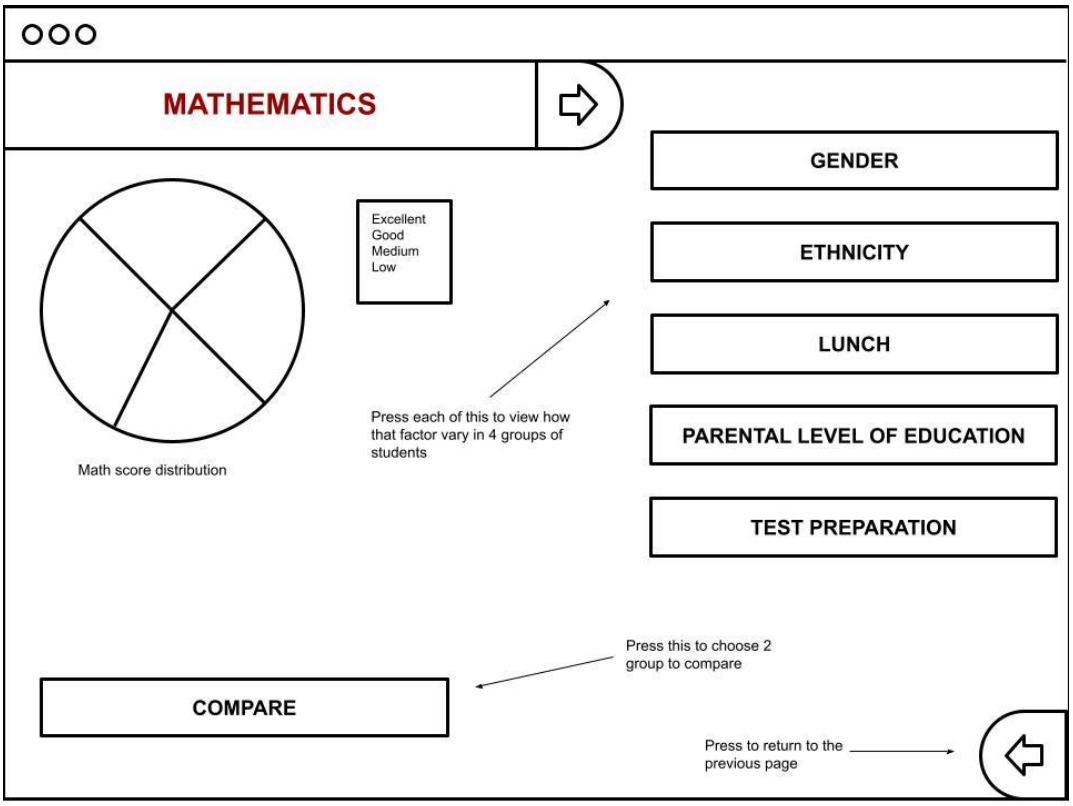
STAGE	ACTION	MEMBER	DEADLINE
PLANNING	Decide the topic and main dataset	All	02/03/2022
	Make the proposal for the project	All	10/03/2022
	Individual topic research	All	20/03/2022
	Decide how to categorize and visualize the data	All	01/04/2022
	Create the a basic website with the main charts and basic interactive feature	All	14/04/2022
	Collect and visualize related dataset	All	25/04/2022
	Add more charts and features to the web	All	15/05/2022

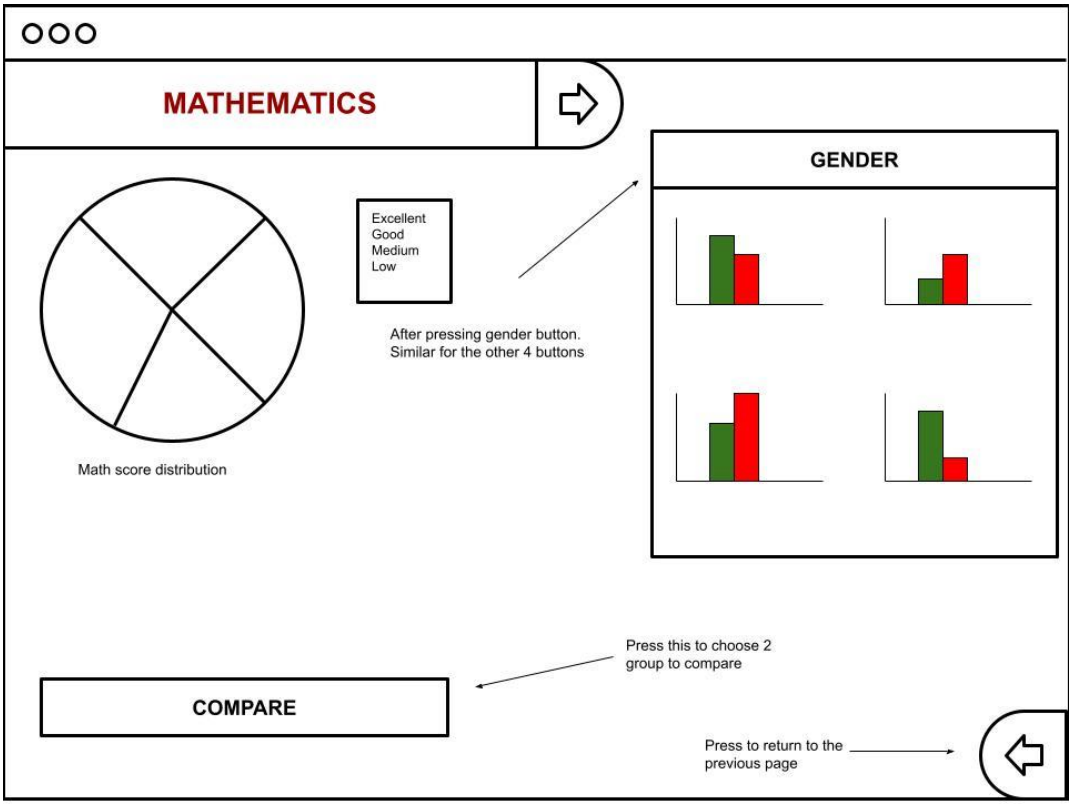
	Finish up the report and the final website	All	20/05/2022
	Prepare for the final demo and presentation	All	23/05/2022
PROCESSING	Processing the data	Khoa	
	Find dataset related to the factors in the main dataset	Phong, Duy & Minh	
ANALYZING & VISUALIZING	Categorize scores and factors into groups	Minh	
	Find the patterns and choose appropriate chart to use	All	
WEBSITE DESIGNING	Graph the charts on the website using data	Duy & Phong	
	Add interactive features to the website	All	
	Design the website and chart's appearance for better visualization	Minh & Khoa	
PRESENTATION	Make the report	All	
	Prepare the slides and demo for the presentation	Minh & Khoa	
	Prepare the script for the presentation	Phong & Duy	

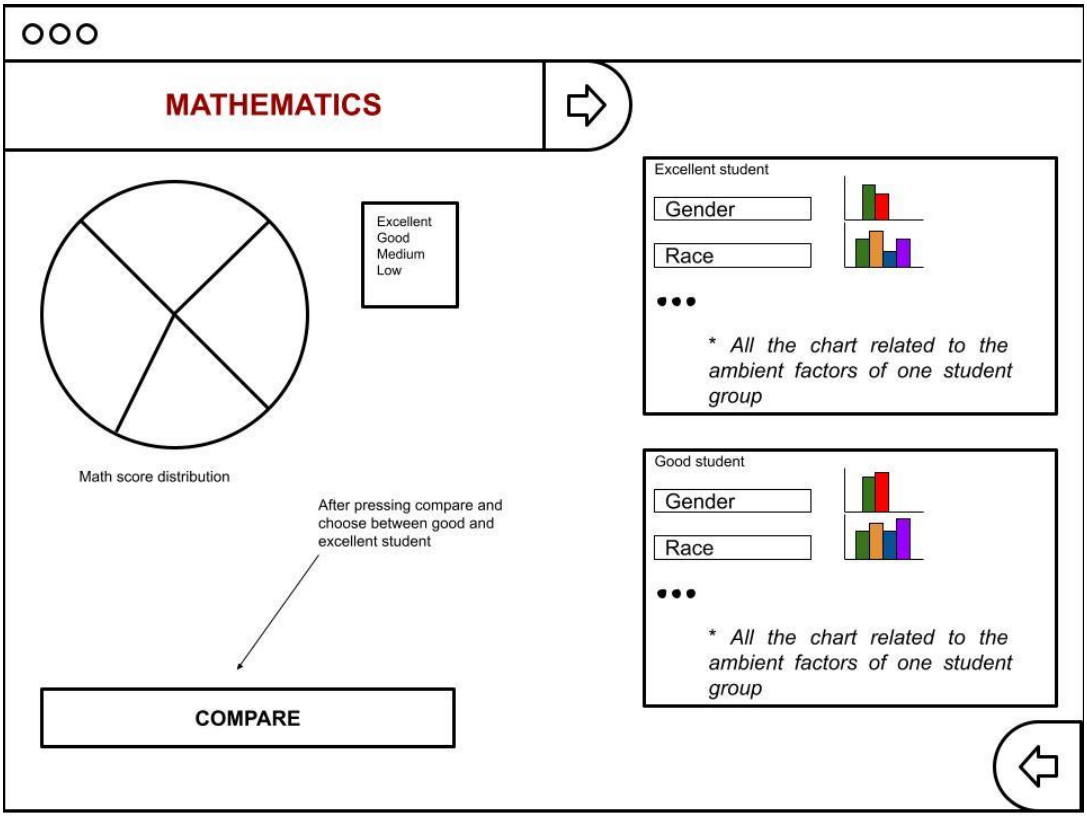
V. PROTOTYPE SKETCHES






MATHEMATICS
GENDER
ETHNICITY
LUNCH
PARENTAL LEVEL OF EDUCATION
TEST PREPARATION
Excellent
Good
Medium
Low
Math score distribution
Press each of this to view how that factor vary in 4 groups of students
Press this to choose 2 group to compare
COMPARE
Press to return to the previous page


MATHEMATICS
GENDER
Excellent
Good
Medium
Low
Math score distribution
After pressing gender button. Similar for the other 4 buttons
Press this to choose 2 group to compare
COMPARE
Press to return to the previous page


MATHEMATICS
Excellent
Good
Medium
Low
Math score distribution
After pressing compare and choose between good and excellent student
COMPARE
Excellent student
Gender
Race
* All the chart related to the ambient factors of one student group
Good student
Gender
Race
* All the chart related to the ambient factors of one student group