MAYBE THERE IS A SHARK IN THE WATER

SOPHIYA SIDDIQUI, MILENKO ROSIĆ,OĞUZHAN ŞAHINGÖZ, DEIMANTE PIRAITYTE

DATA CLEANING

Used data cleaning tecniques in accordance to cleaning process

- **REMOVING UNNECESSARY COLUMNS:**

COLUMNS LIKE 'UNNAMED: 11', 'HREF', 'CASE NUMBER.1', 'UNNAMED: 21', 'UNNAMED: 22', AND 'ORIGINAL ORDER' WERE DROPPED AS THEY WERE EITHER UNNAMED, CONTAINED URLS, OR LACKED INFORMATIVE DATA.

- **RENAMING COLUMNS:**

EMPTY OR UNCLEAR COLUMN NAMES WERE REPLACED WITH MORE DESCRIPTIVE NAMES

- **RENAMING VALUES AND CLEANING TYPOS**

- **DROPPING NULL VALUES:**

ROWS WITH NULL VALUES WERE DROPPED FROM THE DATAFRAME USING THE DROPNA() FUNCTION.

- **ADJUSTING DATA TYPES:**

DATA WRANGLING AND CLEANING

- **Processing huge dataframe.**
- **Handling of missing and misspelled data.**

  Mostly solved by going through unique values and renaming it acordingly.
- **Inconsistent data formats and entries.**

  Columns like State/Country/Location were sharing same entries, mixed with a lot of desriptions rather than entries in columns like “Name”, “Activity”,”Injury"
- **Transparency Issues**

PROJECT OVERVIEW

Problem statement: Shark attack reports are increasing over the past years

Hypothesis: An escalation in oceanic activities, such as surfing and swimming, will correlate with an increase in shark attacks in the USA.

Columns

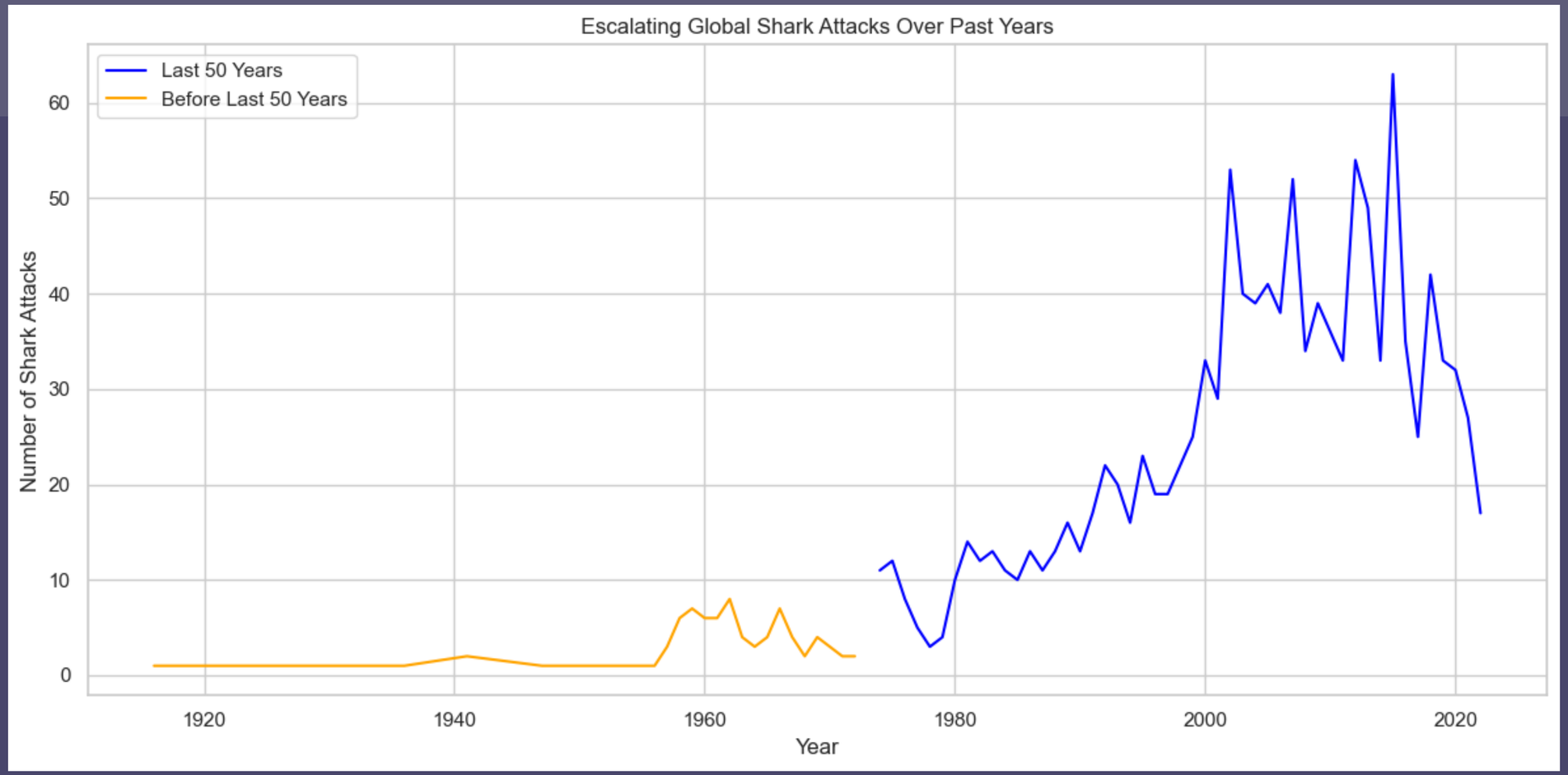
23

Students

4

Rows

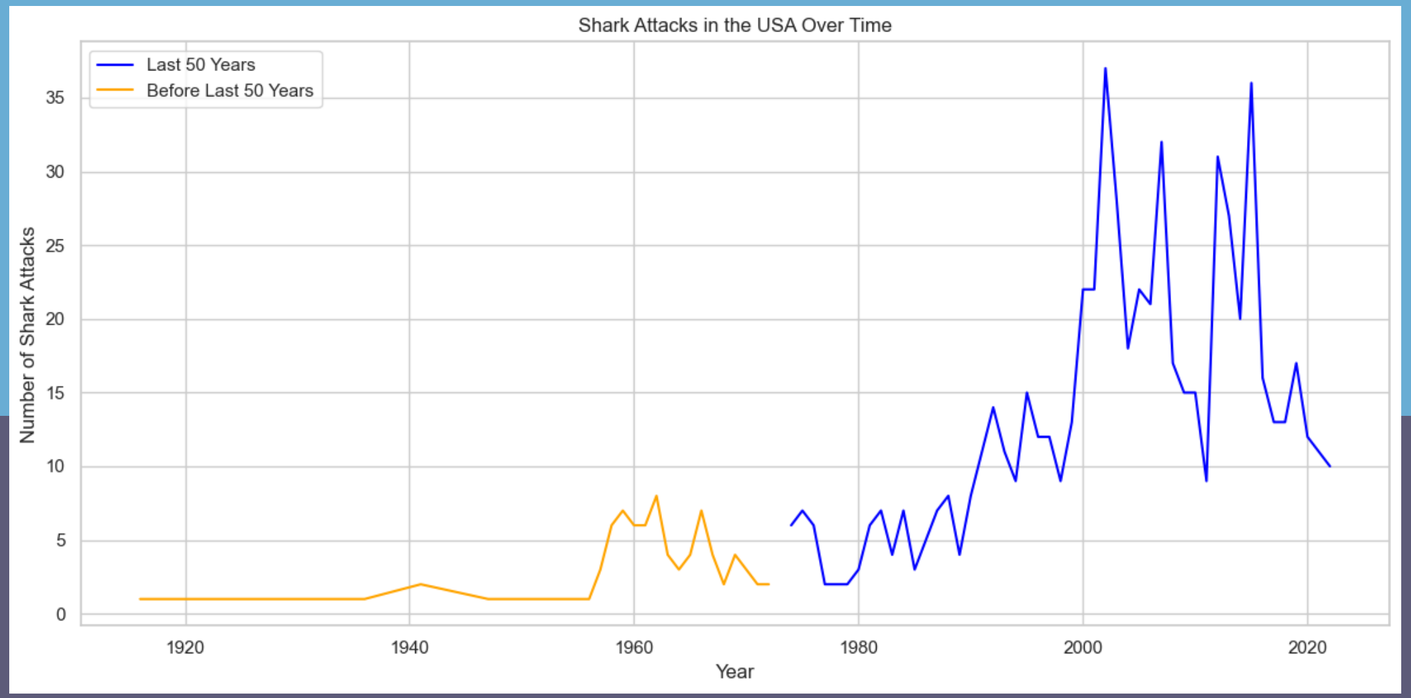
6965

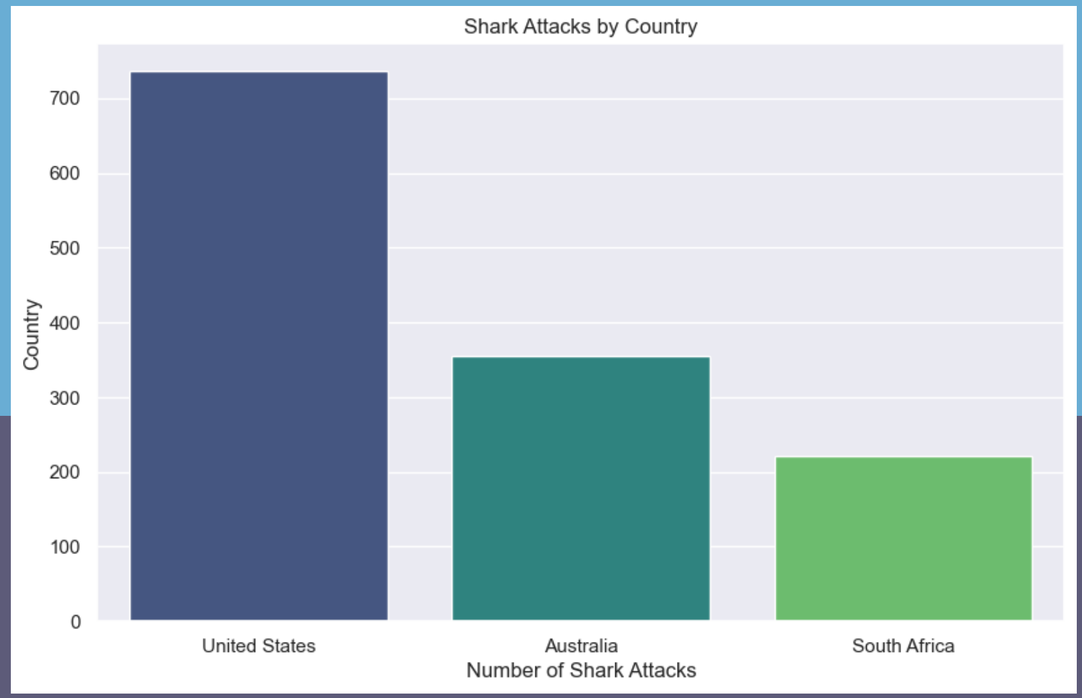


EXPLORATORY DATA ANALYSIS

Used Methods :

- Exploring relationships between variables.
- Performing Hypothesis testing
- Visualising data distibutions

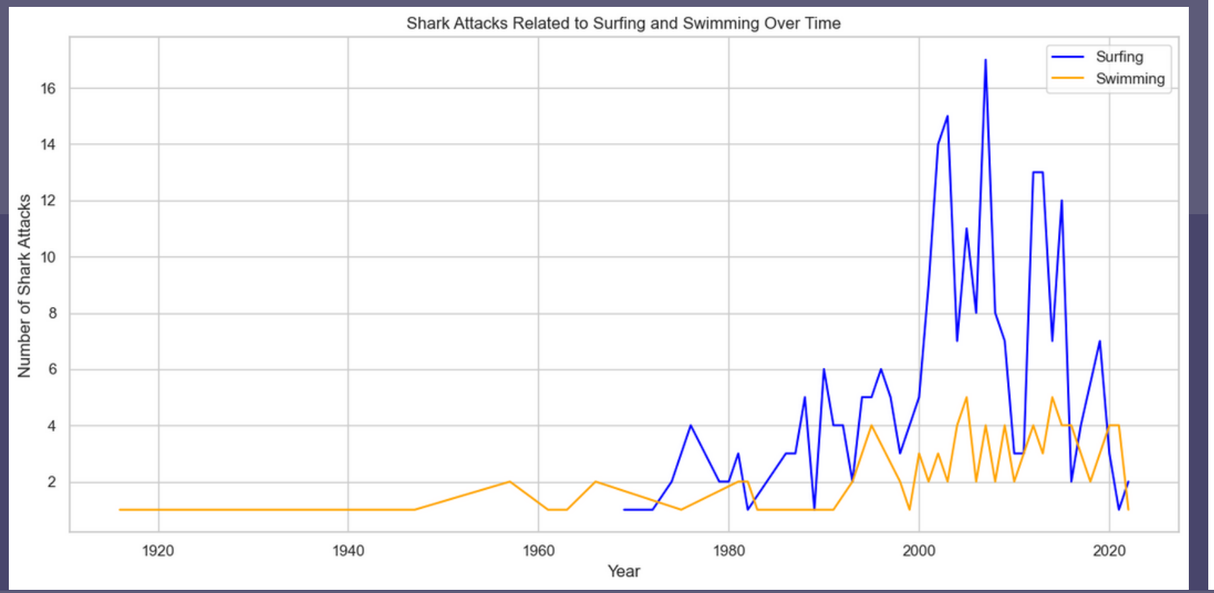


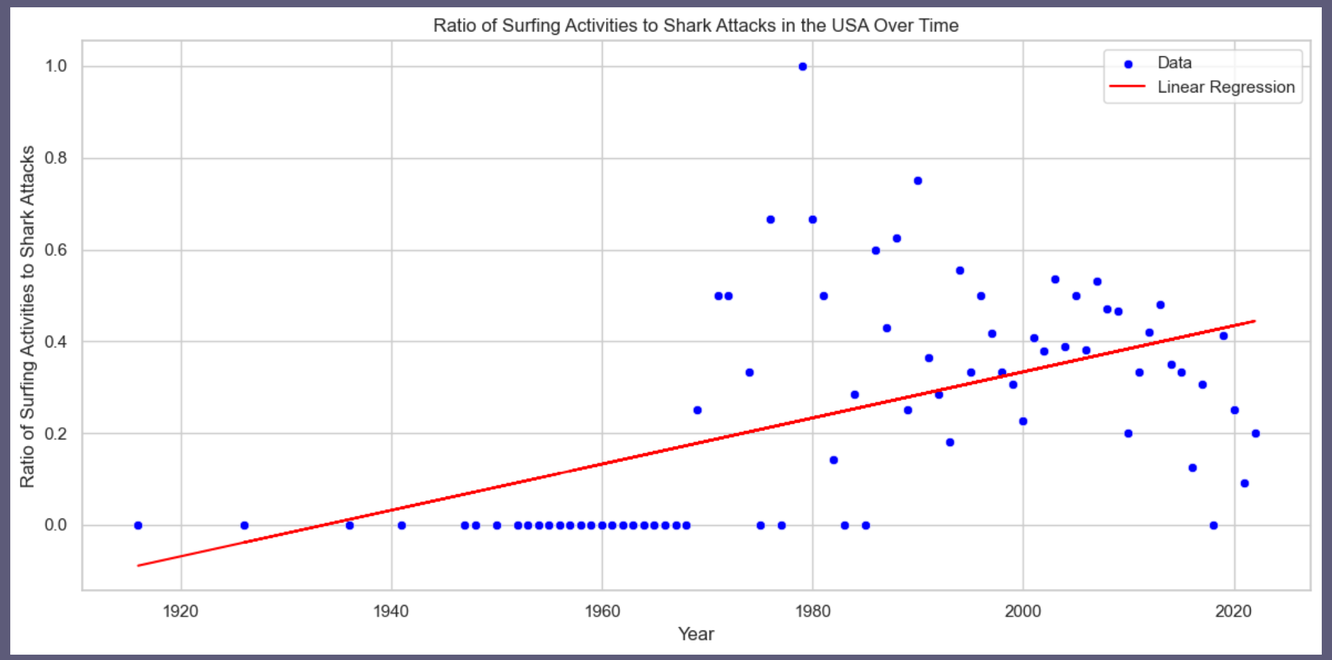


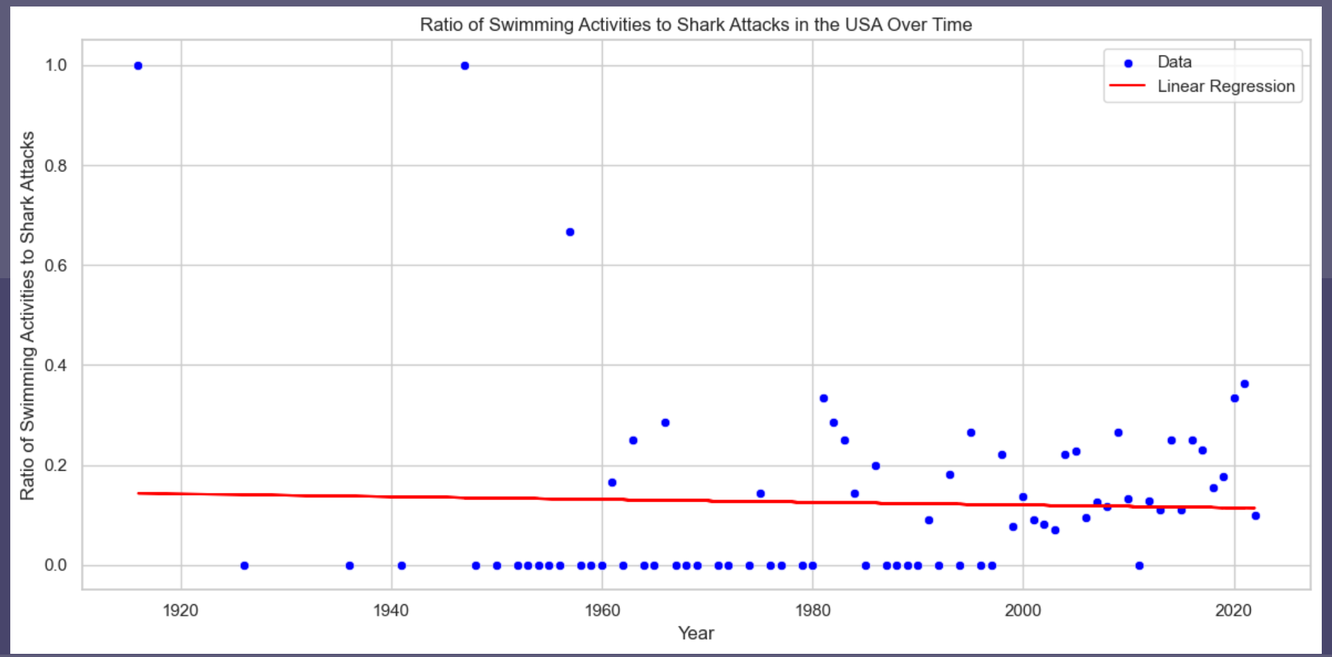
CONCLUSION:

Problem statement:
Increase in shark attack reports in recent years.

Hypothesis was partially accepted as shark attack increase in USA, but only show a moderate correlation to surfing and not other activities such swimming.







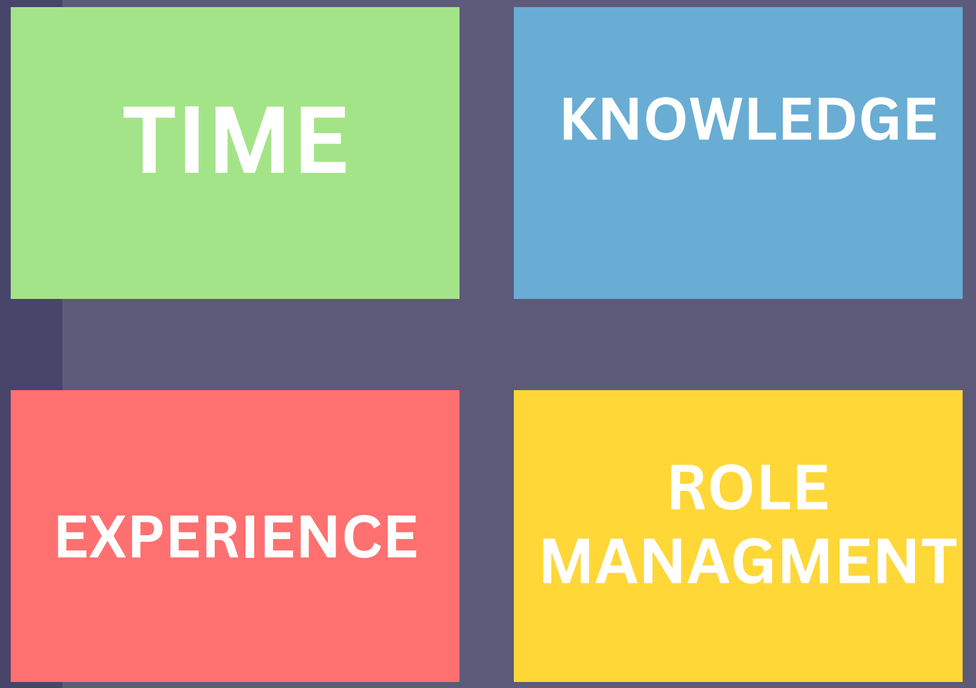
MAJOR OBSTACLE

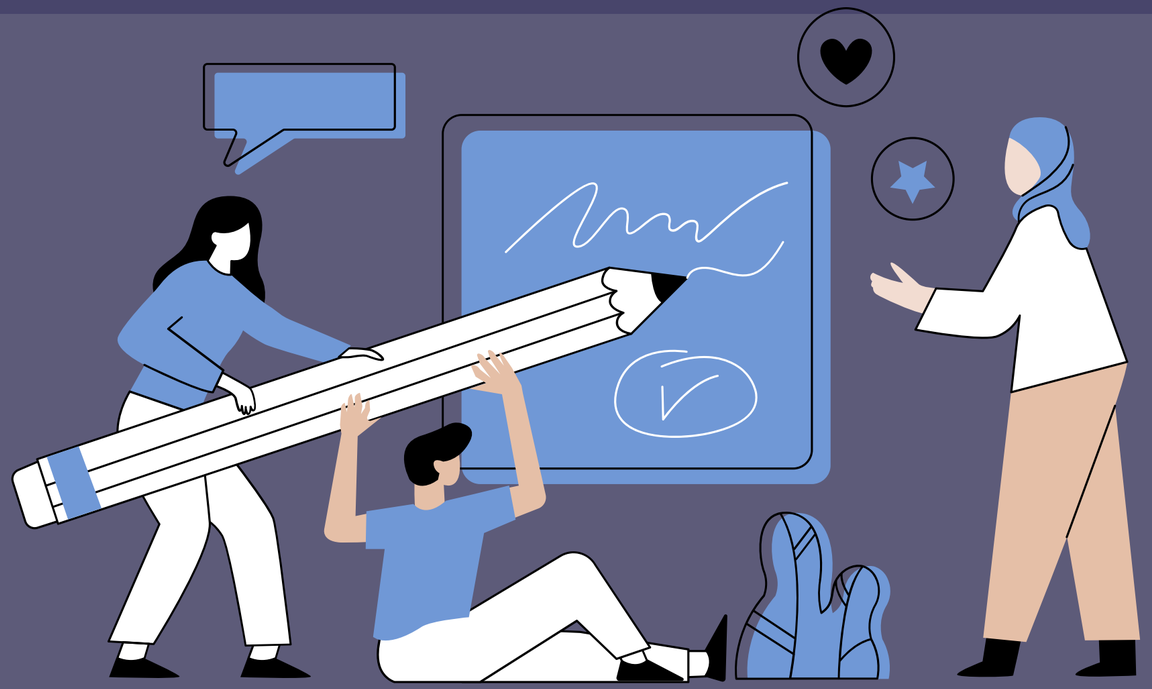
01

02

03

- Limited time exploring extra features and improvements.
- Very little to no experience with cleaning stressed out our decision making process.
- Comparing to a day 1 we gained a lot of insightful experience that would have allowed us to make some things in more efficient manner.





THANK YOU,BYE

Milenko Rosić

HVALA TI

Oğuzhan Şahingöz,

TEŞEKKÜRLER

Sophiya Siddiqui

धन्यवाद

Deimante Piraityte

AČIŪ