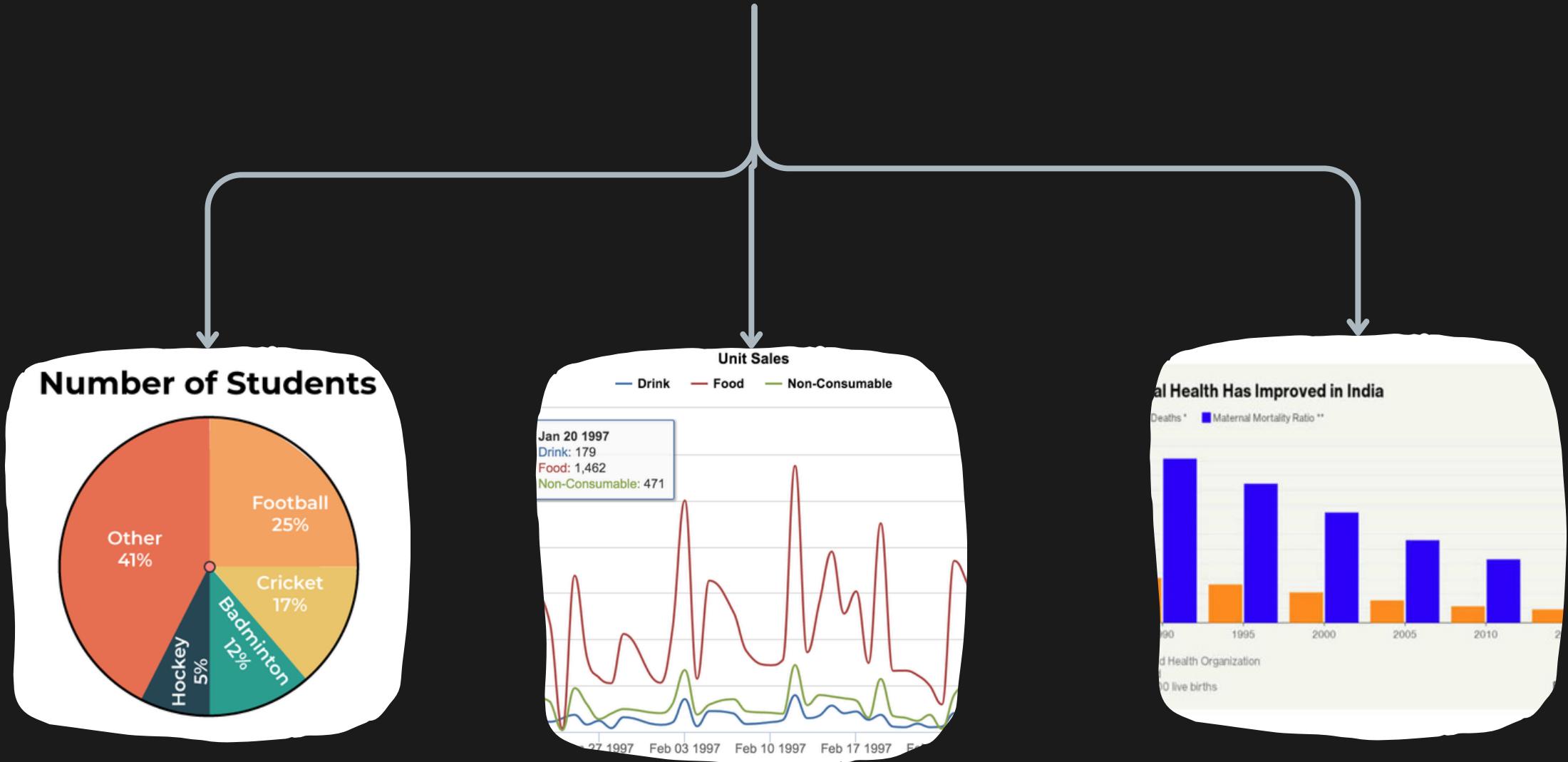


AI CHART CAPTIONING

AI empowers us to transform possibilities into reality

Lithikhaa Kathiresan
KGISL Institute of Technology

VISUALIZATION OF CHARTS



Pie Chart

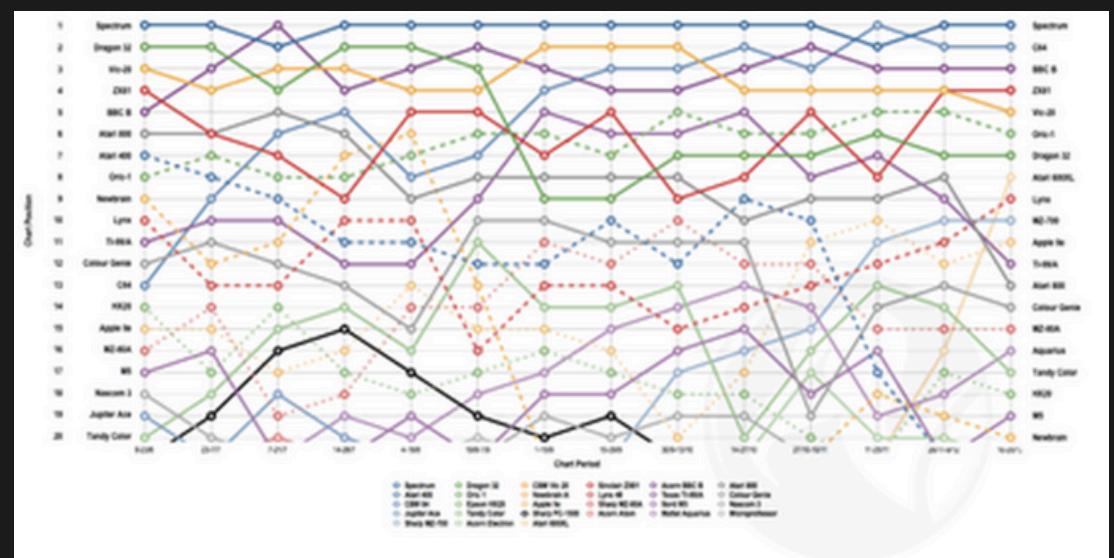
Area Chart

Bar Chart

Why do we need a
Chart?

Problem Statement

Modern charts are increasingly complex, **making it difficult for common users** to interpret and understand visual data effectively

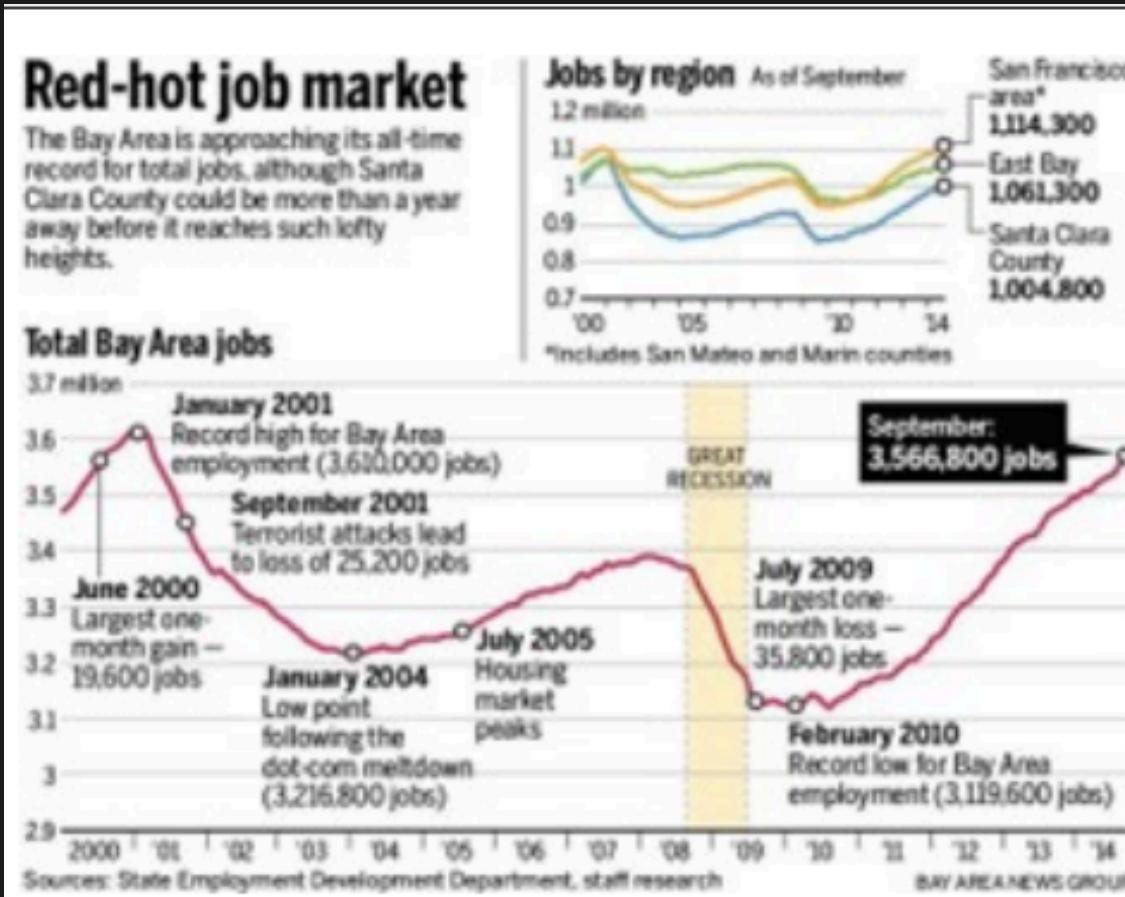


Existing System

Current models overly focus on **data tables, neglect visual patterns**, and struggle with new or diverse chart types in real-world situations.

Region	Product A	Product B
North	200	150
South	250	300
East	100	200
West	300	250

Chart Image



Summary Generated from the Data Table

The provided data table shows the total number of jobs in the Bay Area, as well as the number of jobs in San Francisco, the East Bay, and Santa Clara County, from 2000 to 2026. Unfortunately, the table only provides data for the year 2000, with the remaining years missing data for all regions except the total Bay Area jobs. While we can see the total Bay Area job growth from 3.3 million in 2000 to a projected 4.0 million in 2026, we lack the specific data to assess job trends within San Francisco, East Bay, and Santa Clara County over the years. To draw meaningful conclusions about job market trends in these regions, it is necessary to have complete data for all years and all regions.

How can we improve AI models ? to better interpret diverse chart visuals?

Proposed Solution

ChartVision learns directly from chart images, capturing important visual details and trends that might be missed by data tables.

01

Summarizes the Chart

02

Answers Questions

03

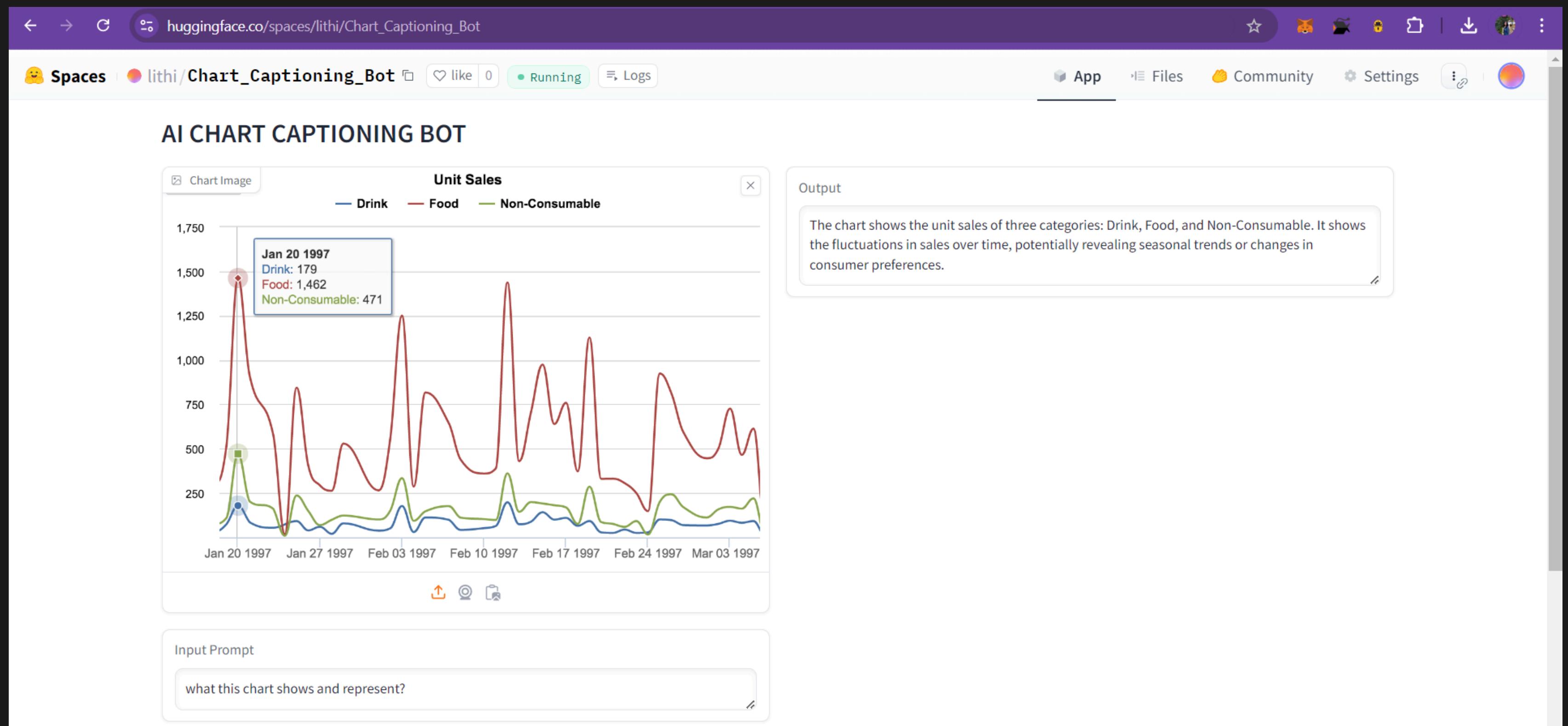
Fact-Checking

04

Multi Language

01

Summarizes the Chart



Answers Questions

huggingface.co/spaces/lithi/Chart_Captioning_Bot

AI CHART CAPTIONING BOT

Chart Image

Number of Students

The chart displays the following data:

Sport	Percentage
Football	25%
Cricket	17%
Badminton	12%
Hockey	5%
Other	41%

Output

The highest value is 41% and the lowest value is 5%.

Input Prompt

what is the highest value and lowest value in graph/chart?

Answers Questions

Spaces | lithi/Chart_Captioning_Bot | like 0 | Running | Logs | App | Files | Community | Settings |

AI CHART CAPTIONING BOT

Chart Image

Many think China's global influence will decline after the coronavirus outbreak

% who say each will have ___ influence in world affairs after the coronavirus outbreak compared to before the outbreak

	More	About the same	Less
U.S.	29%	41%	29%
EU	19	59	21
China	17	31	50

Note: No answer responses not shown.
Source: Survey of U.S. adults conducted April 29-May 5, 2020.
"Americans Give Higher Ratings to South Korea and Germany Than U.S. for Dealing With Coronavirus"

Output
 $(19+31+17)/3=23.33$

Input Prompt
What's the average of all the values in the green bars (round to one decimal)?

Answers Questions

Spaces | lithi/Chart_Captioning_Bot | like 0 | Running | Logs | App | Files | Community | Settings | ⚙️ | 🎨

AI CHART CAPTIONING BOT

Chart Image

Age Group	Facebook Messenger (%)	WhatsApp (%)
18-29	73%	30%
30-59	66%	25%
60+	43%	6%

Share of respondents

Legend: ● Facebook Messenger ● WhatsApp

© Statista 2021 | Show source

Output

Age	Facebook Messenger	WhatsApp
18-29	73%	30%
30-59	66%	25%
60+	43%	6%

Input Prompt

convert above chart content into tabular format?

Fact-Checking

Spaces | lithi/Chart_Captioning_Bot | like 0 | Running | Logs | App | Files | Community | Settings | ⚙️ | 🌐

AI CHART CAPTIONING BOT

Chart Image Index (RPI) inflation forecast United Kingdom (UK) 2019 to 2040

Year	RPI inflation
2020	3.2
2021	3.1
2022	3.3
2023	3.6
2024	3.7
2025	4.0
2026	4.1
2027	4.2

Output
True.

Input Prompt
is that statement is true in 20240 the RPI inflation increases?

Multi Language

Spaces | lithi/Chart_Captioning_Bot | like 0 | Running | Logs | App | Files | Community | Settings | ⚙️ | 🌐

AI CHART CAPTIONING BOT

Chart Image

Age Group	Facebook Messenger (%)	WhatsApp (%)
18-29	73%	30%
30-59	66%	25%
60+	43%	6%

Share of respondents

© Statista 2021 | Show source

Additional Information

Input Prompt

விளக்கப்படத்துன் வகை என்ன

The chart is a bar chart showing the share of respondents who use Facebook Messenger and WhatsApp by age group.

Multi Language

Spaces | lithi/Chart_Captioning_Bot | like 0 | Running | Logs

App Files Community Settings

AI CHART CAPTIONING BOT

Chart Image

A bar chart titled "Share of respondents" comparing the usage of Facebook Messenger (blue bars) and WhatsApp (dark blue/black bars) across three age groups: 18-29, 30-59, and 60+. The y-axis represents the percentage from 0% to 100% in increments of 20%. Facebook Messenger usage is highest in the 18-29 age group at 73%, followed by 66% for 30-59, and 43% for 60+. WhatsApp usage is 30% for 18-29, 25% for 30-59, and 6% for 60+.

Age Group	Facebook Messenger (%)	WhatsApp (%)
18-29	73%	30%
30-59	66%	25%
60+	43%	6%

Output

Bar Chart

Share of respondents

© Statista 2021

Additional Information Show source

Input Prompt

quel est le type de graphique

Demo

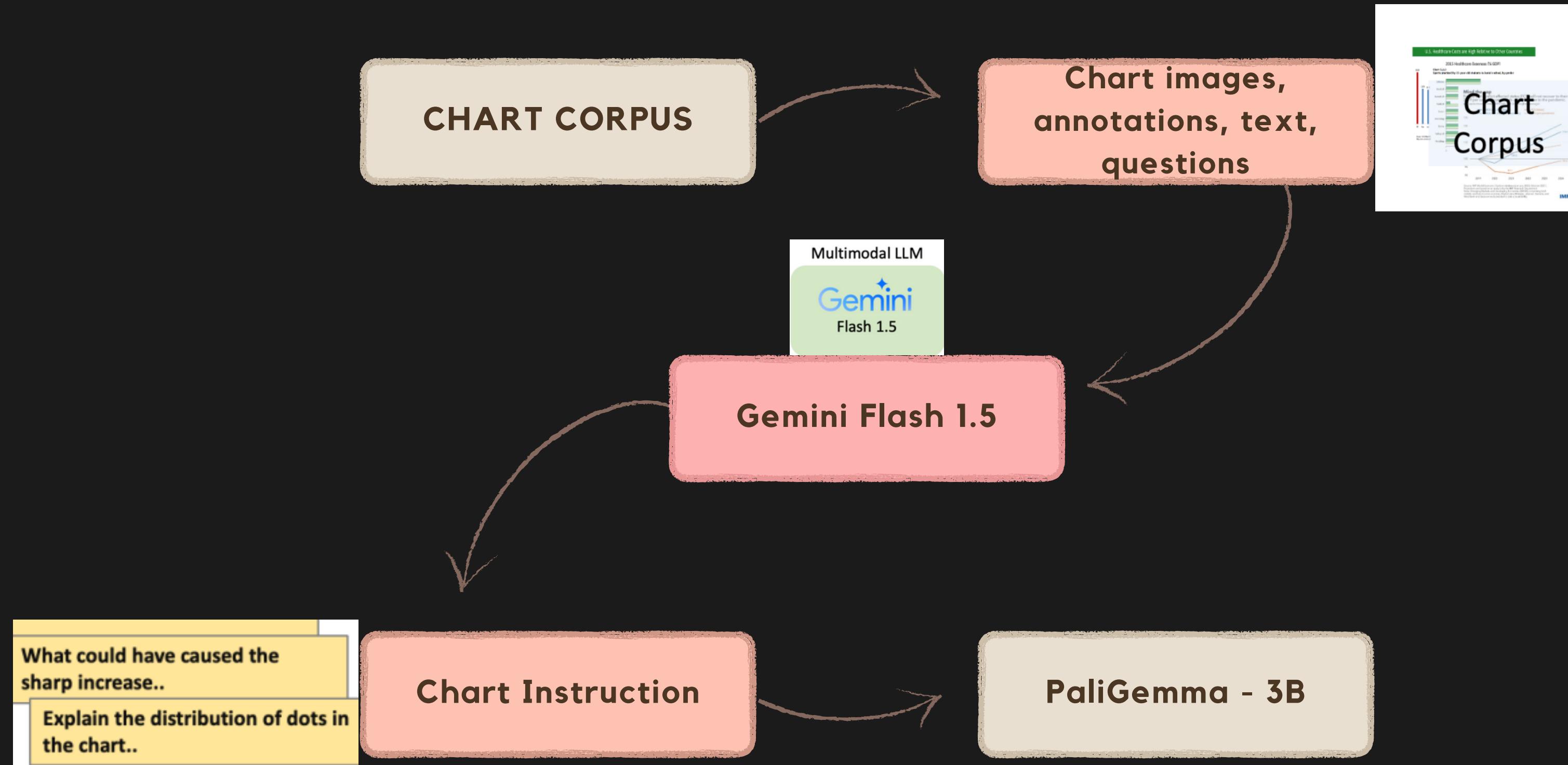
.....

Deployed in HuggingFace



Hugging Face

How It Works



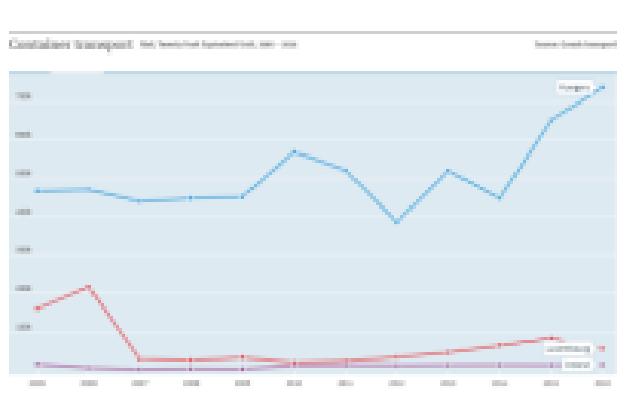
With a refined understanding of chart interpretation, **Paligemma** can now generate relevant insights based on **new chart inputs**

The Model can Identify ?

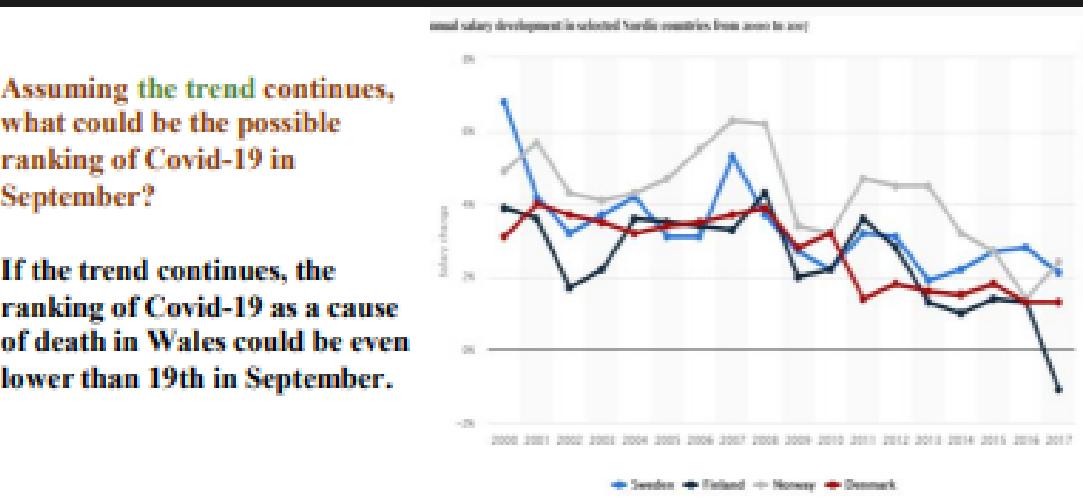
ANOMALY DETECTION

Identify any unusual or unexpected data points in the chart.

The sharp decline in container transport by rail in Hungary from 2010 to 2012 stands out as an unusual point. This is particularly notable considering the overall increasing trend in transport volume for Hungary.'



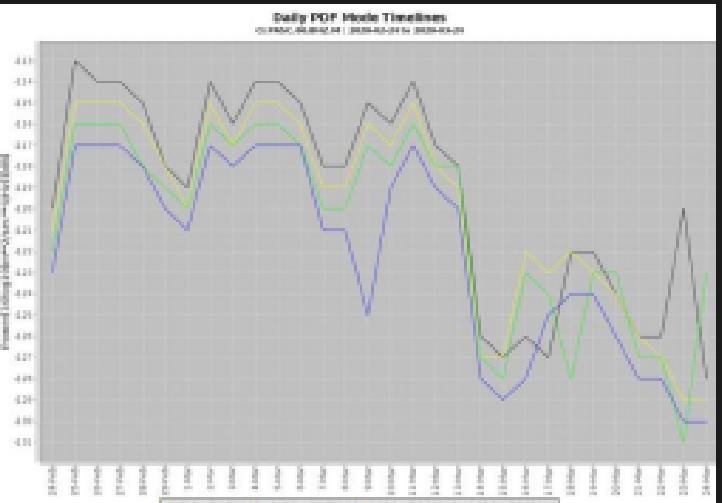
DATA EXTRAPOLATION



TIME SERIES ANALYSIS

Identify the dates when the black line experiences a significant drop in power.

The black line experiences significant drops in power around March 13th and March 21st.



Depth Analysis of Working

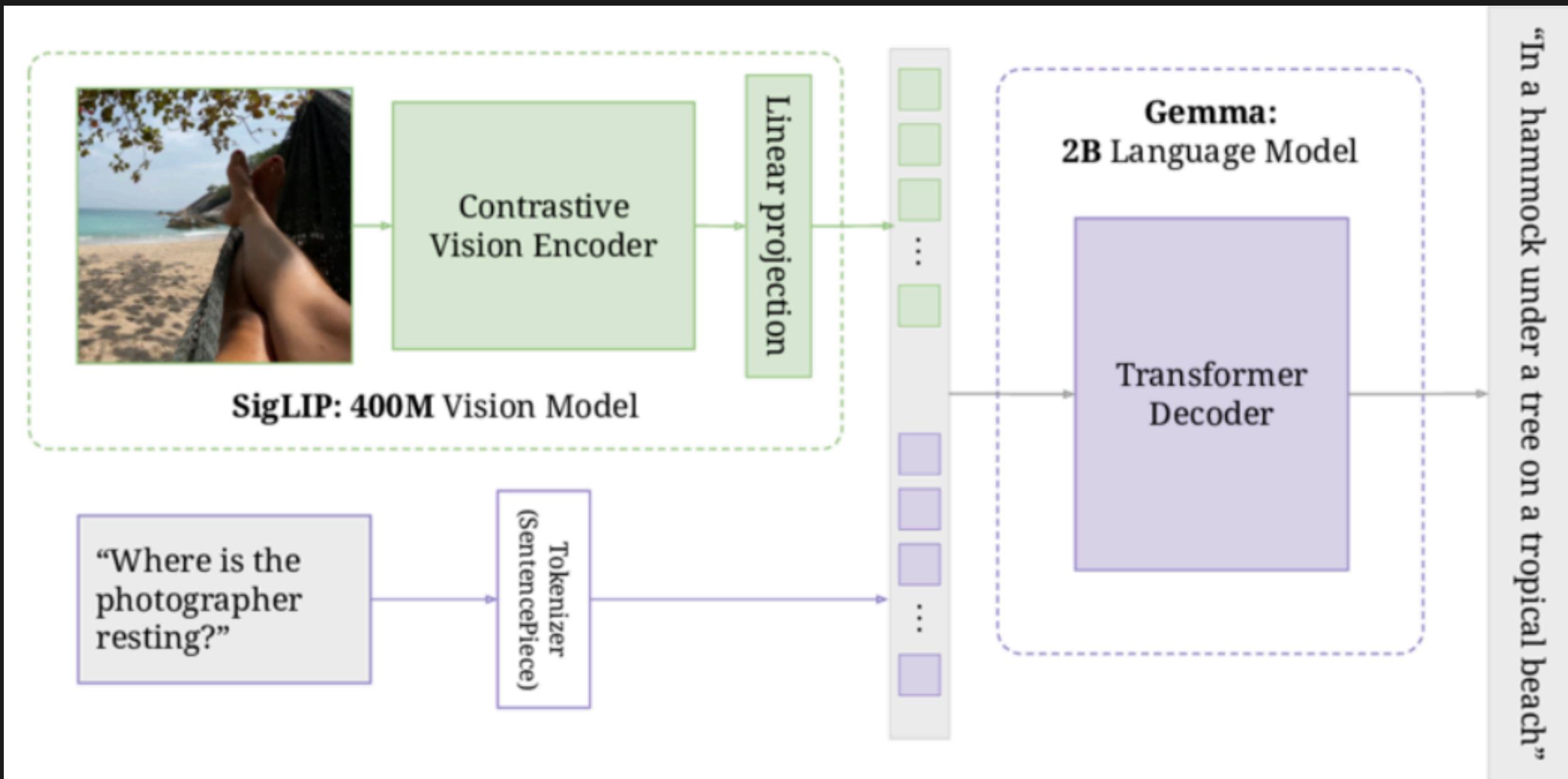


Image Processing

01

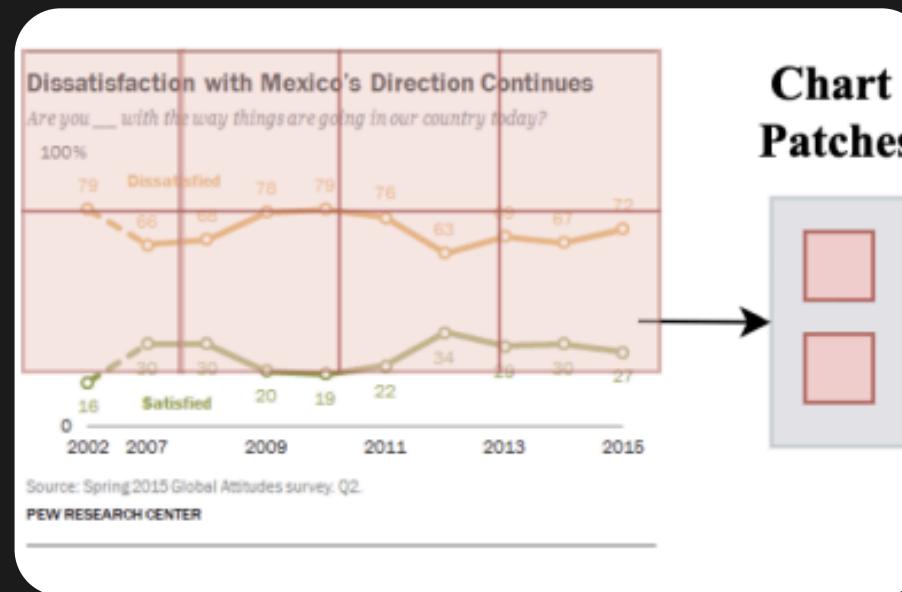


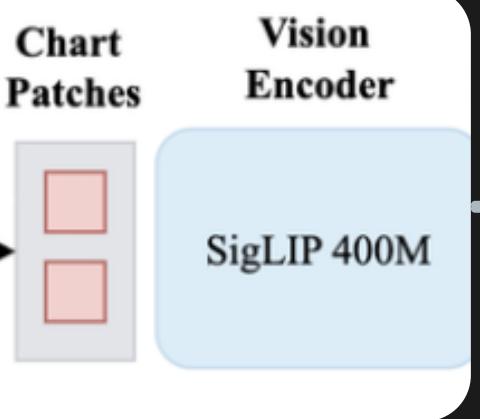
Chart
Patches

A chart image is divided into a grid of patches

For example a chart image with a resolution of 448×448 pixels is split into a 32×32 grid of smaller sections or patches.

Vision Encoding

02



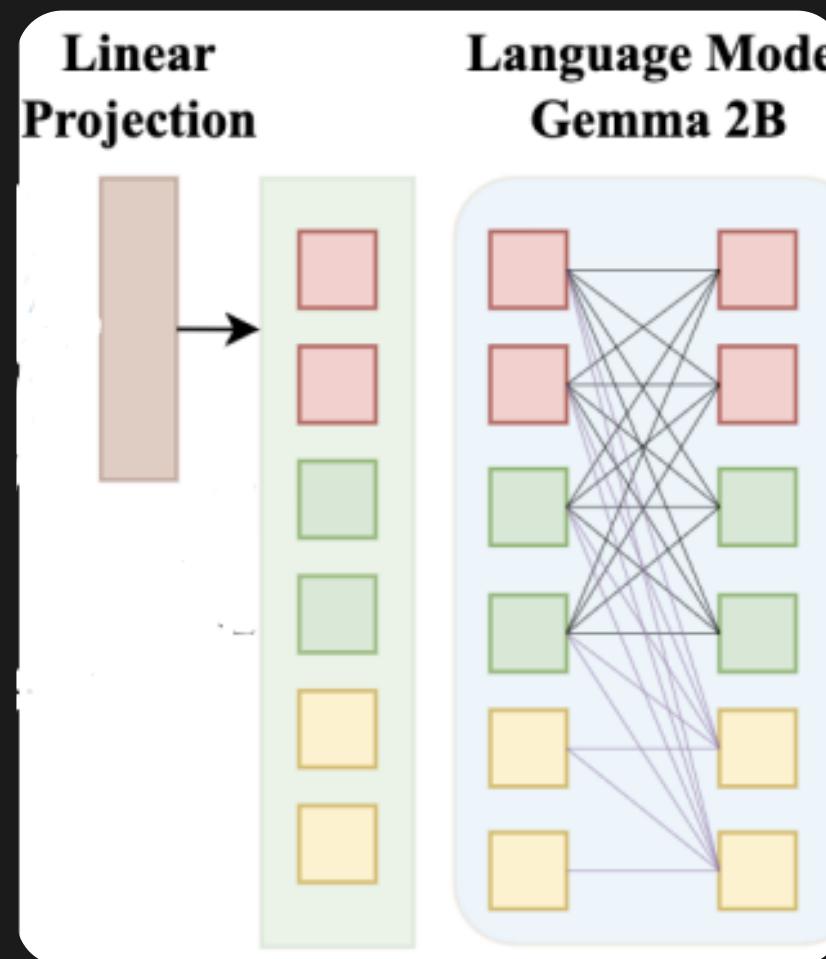
Each image patch is processed separately by the **vision encoder (SigLIP)**

Extract visual features such as lines, shapes, and colours from each patch

The output will be
Numerical representations
(features) of the visual
content from each patch

Text Information Handling

03

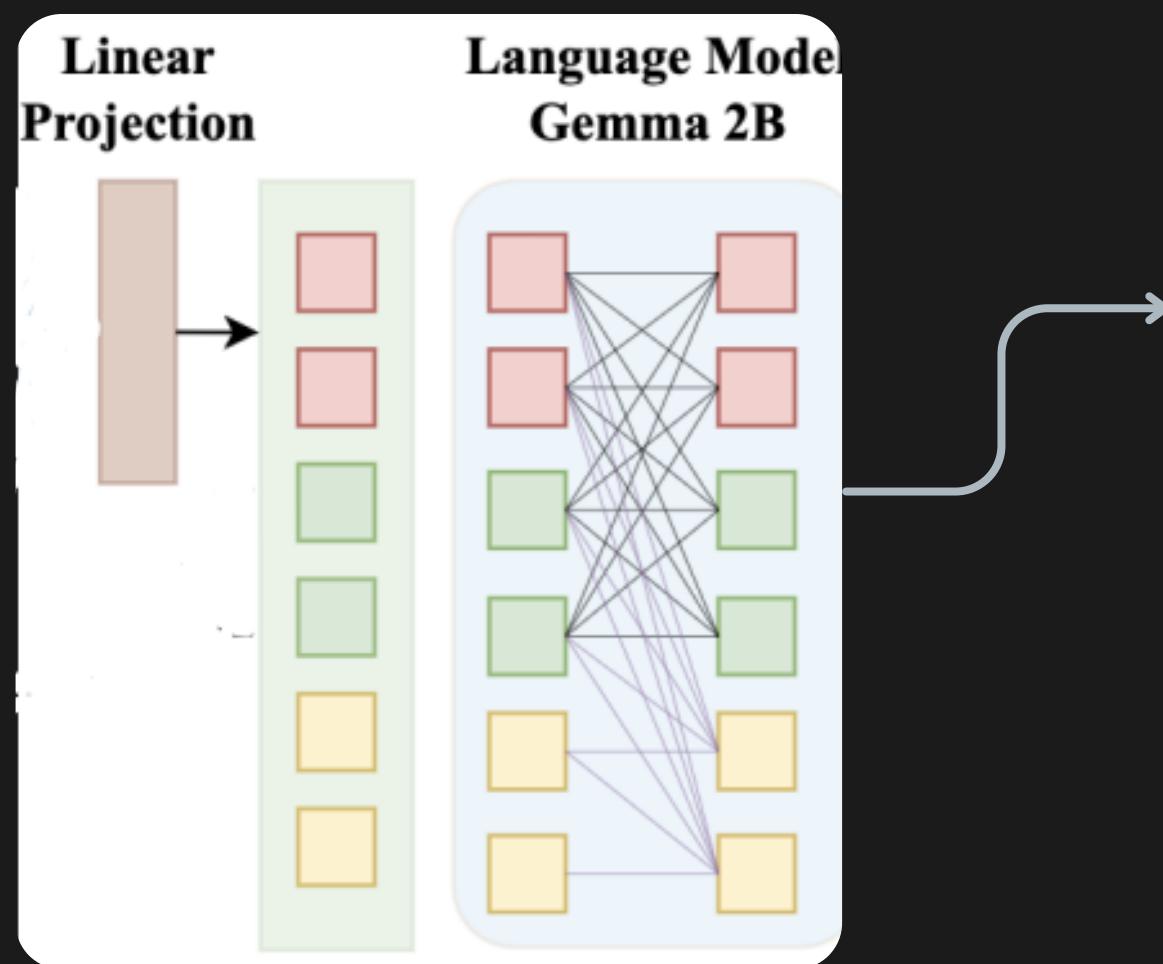


Text related to the chart (like titles, and labels) is converted into **numerical representations**

Visual features from **image patches** are **combined with text embeddings**

Attention Mechanism

04



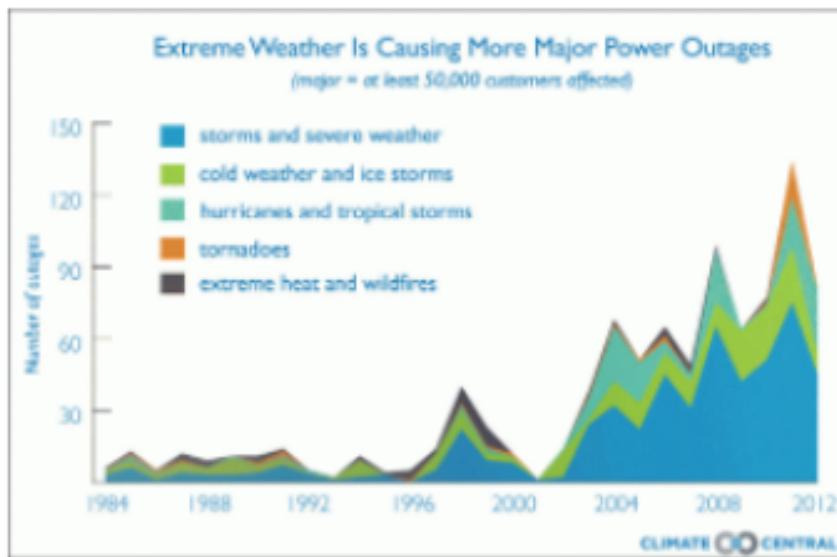
The model uses full attention to analyze **visual features and text Information**, understanding their relationships.

Contextual Understanding Output

05

Compare the number of major power outages caused by hurricanes and tropical storms in 1992 and 2004.

In 1992, the number of major power outages caused by hurricanes and tropical storms was around 10. In 2004, the number was around 60.



The model interprets how the visual and textual data relate to each other provide **accurate descriptions, summaries, or answers related to the chart**

What are our key use cases?



Customer Insights



Data Journalism

Business Reporting

Research Documentation

Market Analysis

Educational Tools

Limitations

Restrictions: A high-performance system is needed for large-scale models.

Image Resolution Limitations: Large images reduce detail and increase processing time.

Closed-Source Dependence: Relying on closed tools affects model reliability and consistency

Model Hallucinations: Incorrect or fabricated information leads to inaccurate or unreliable outputs

Thank You