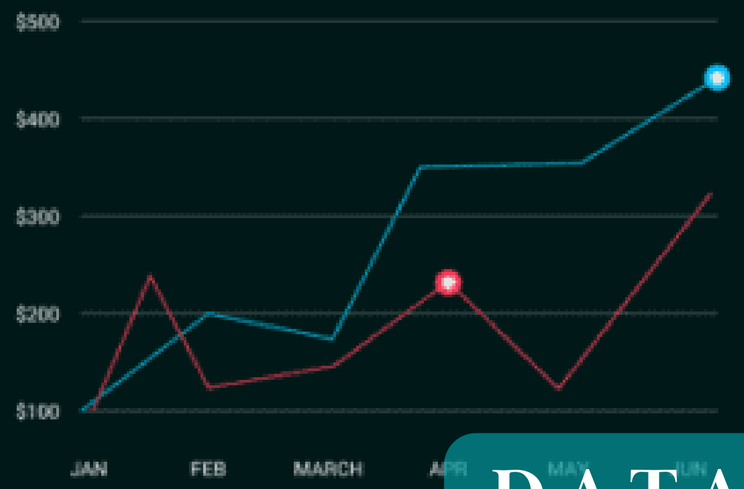
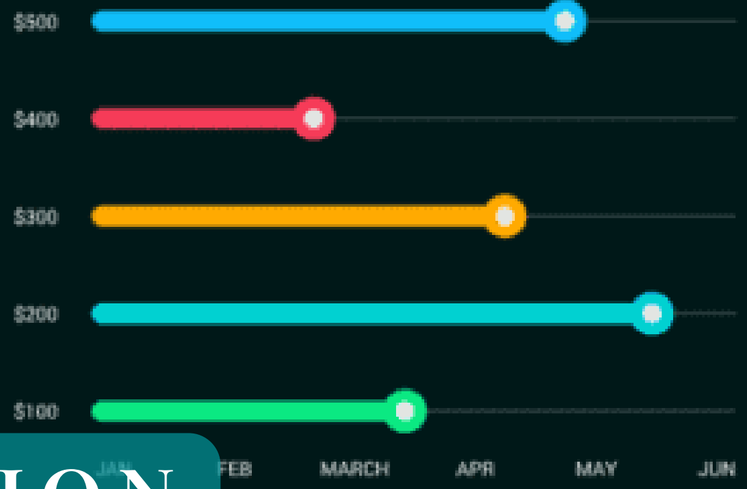
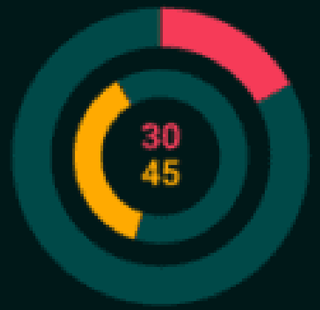


Su	Mo	Tu	We	Th	Fr	Sa
			01	02	03	04
05	06	07	08	09	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	



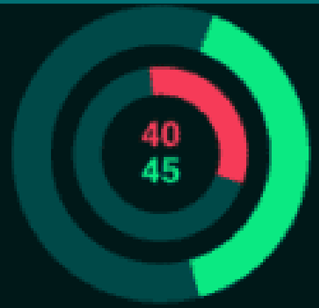
DATA VISUALIZATION



30 70

45 55

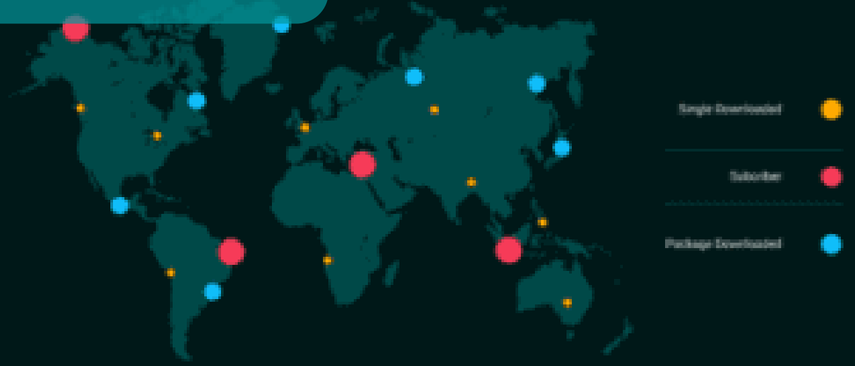
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer ante elit.

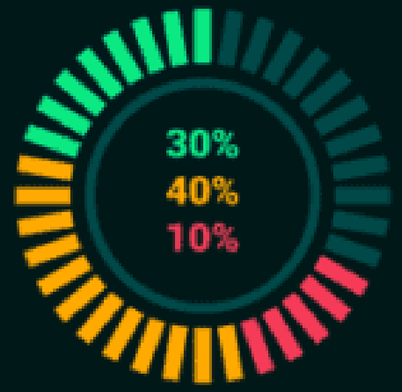


40 60

45 55

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer ante elit.

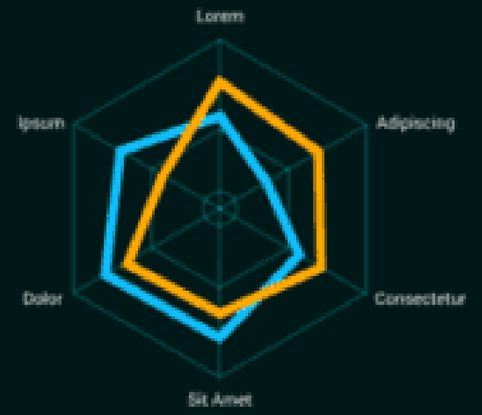




Lorem

Ipsum

Dolor



60%
LOREM IPSUM

75%
LOREM IPSUM

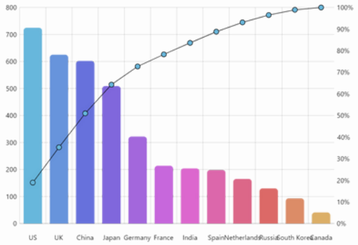
45%
LOREM IPSUM

60%
LOREM IPSUM

How to increase our customer engagement?

Web scraping
Database

Duplicated values,
Structural errors,
Outliers,
Missing data,
Sort and organize data,
irrelevant data,
Validate data,
Transform data



Exploratory
Data Analysis

Define Question | Collect Data | Clean Data | EDA | Report

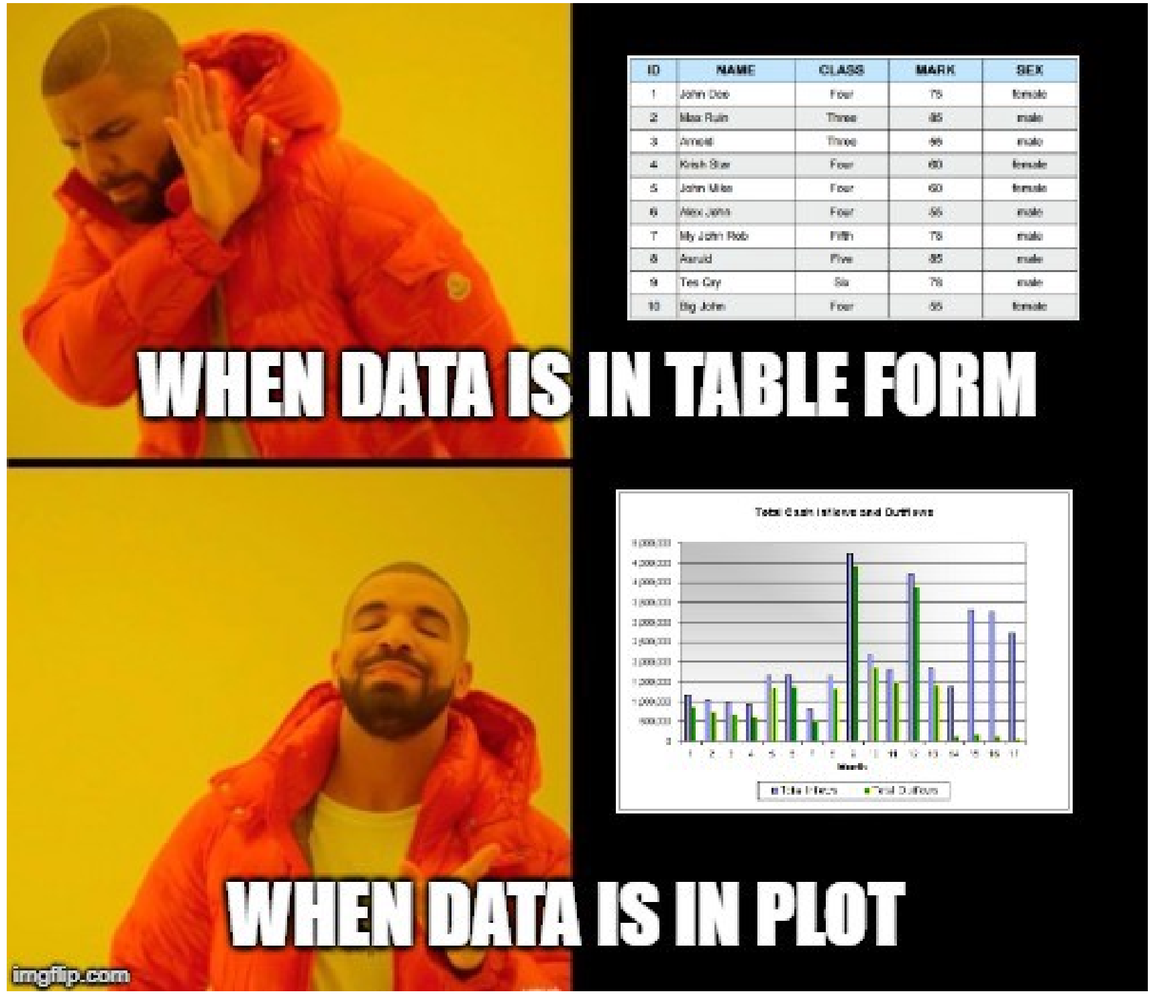
Python,
Oracle,
MySQL,
SQLServer,...

Pandas

Pandas

Matplotlib
Seaborn

PowerBI,
Tableau,
Google Data
Studio

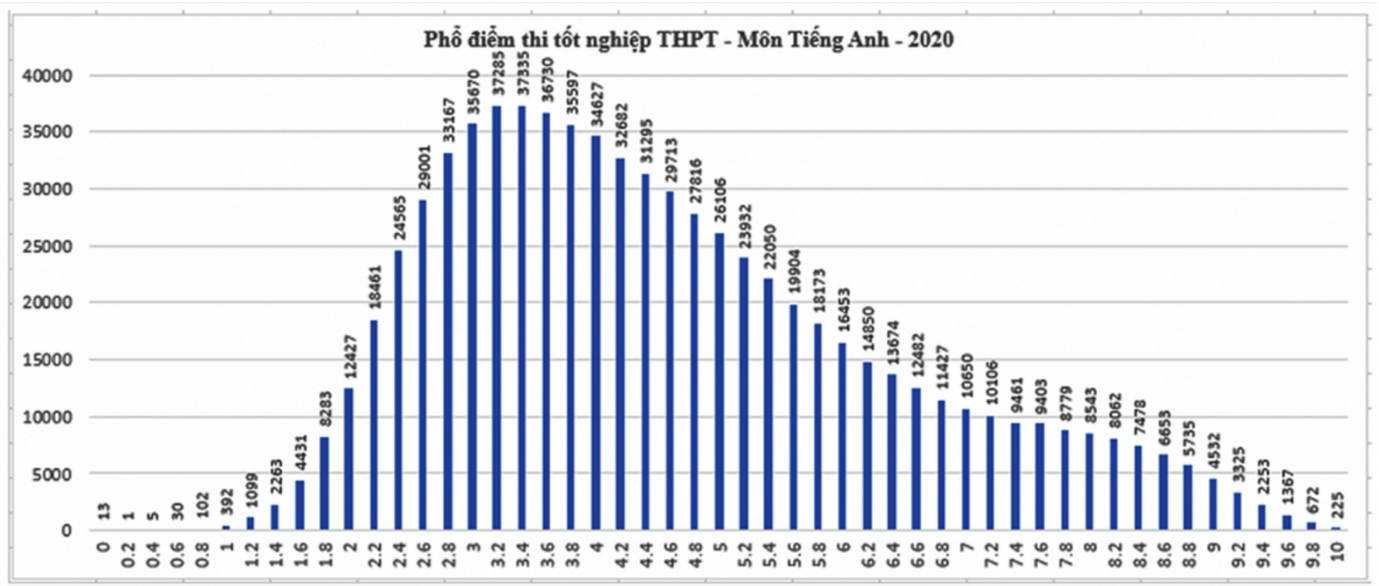

ID NAME CLASS MARK SEX
1 John Doe Four 75 female
2 Max Ruin Three 85 male
3 Arnold Three 90 male
4 Krish Star Four 60 female
5 John Mike Four 60 female
6 Alex John Four 55 male
7 My John Rob Fifth 78 male
8 Asruid Five 85 male
9 Tes Qry Six 78 male
10 Big John Four 55 female
WHEN DATA IS IN TABLE FORM
WHEN DATA IS IN PLOT
imgflip.com

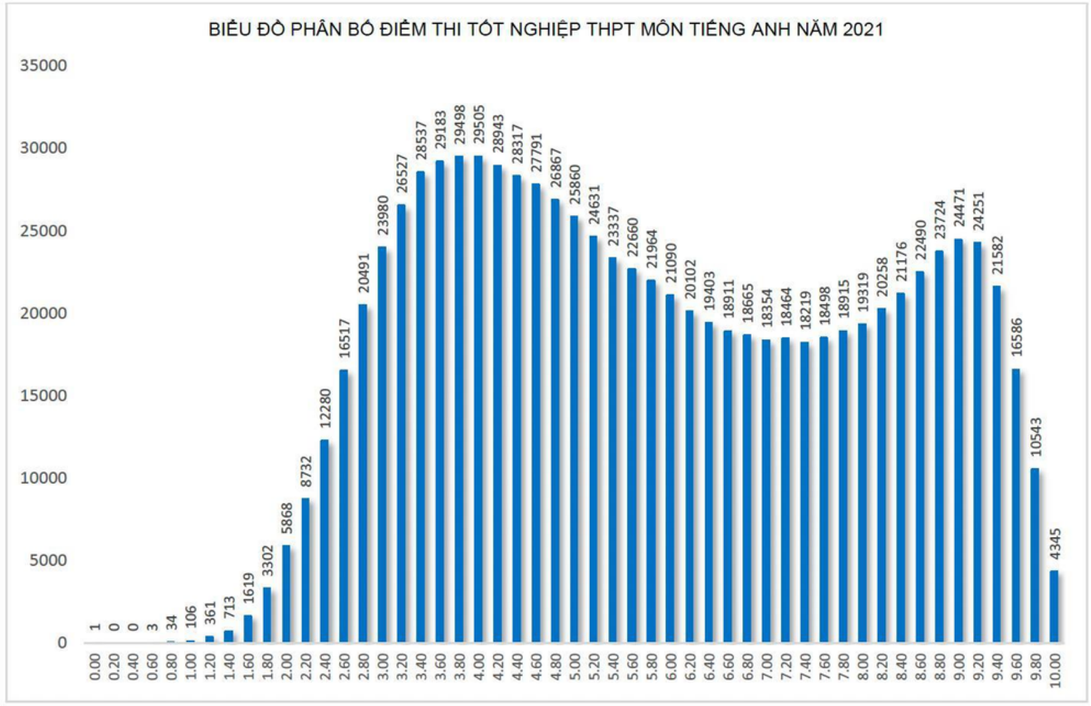
WHY?

WHAT?

HOW?

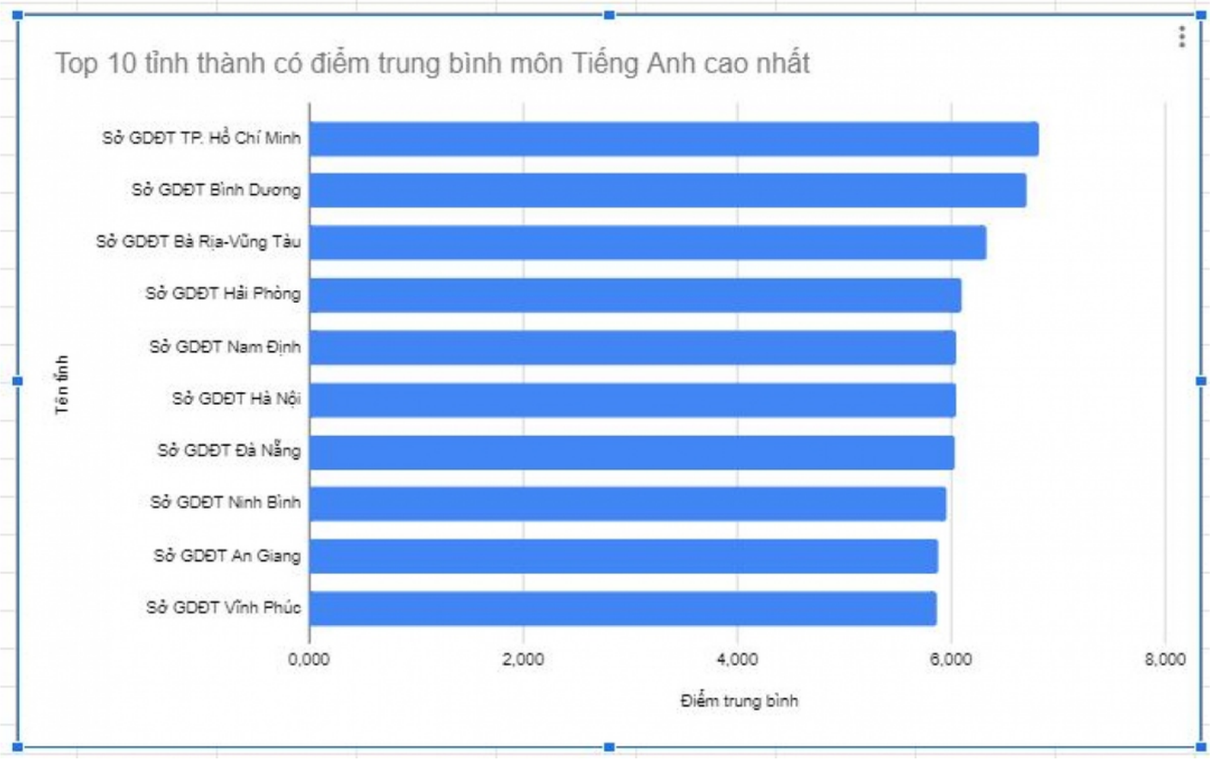
WHY?

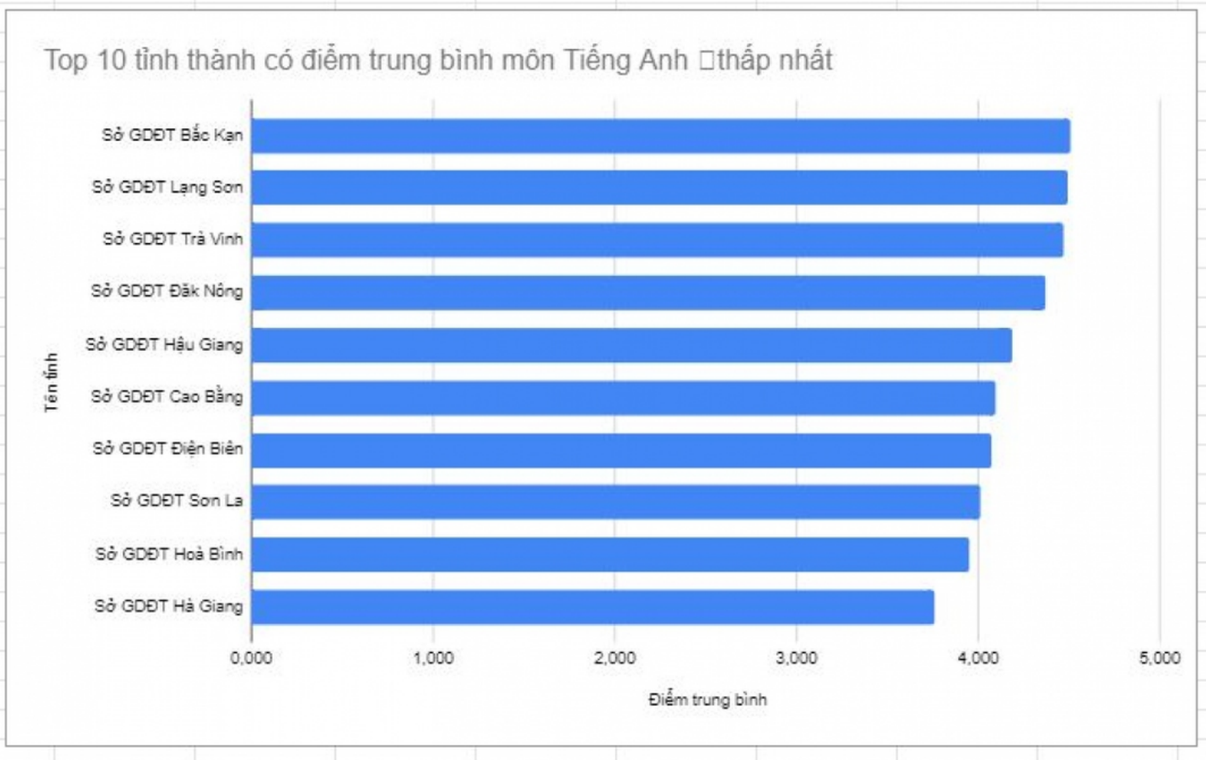




What do you think?

Source: *https://thanhnien.vn/pho-diem-la-nhat-mon-tieng-anh-noi-len-dieu-gi-post1093873.html*

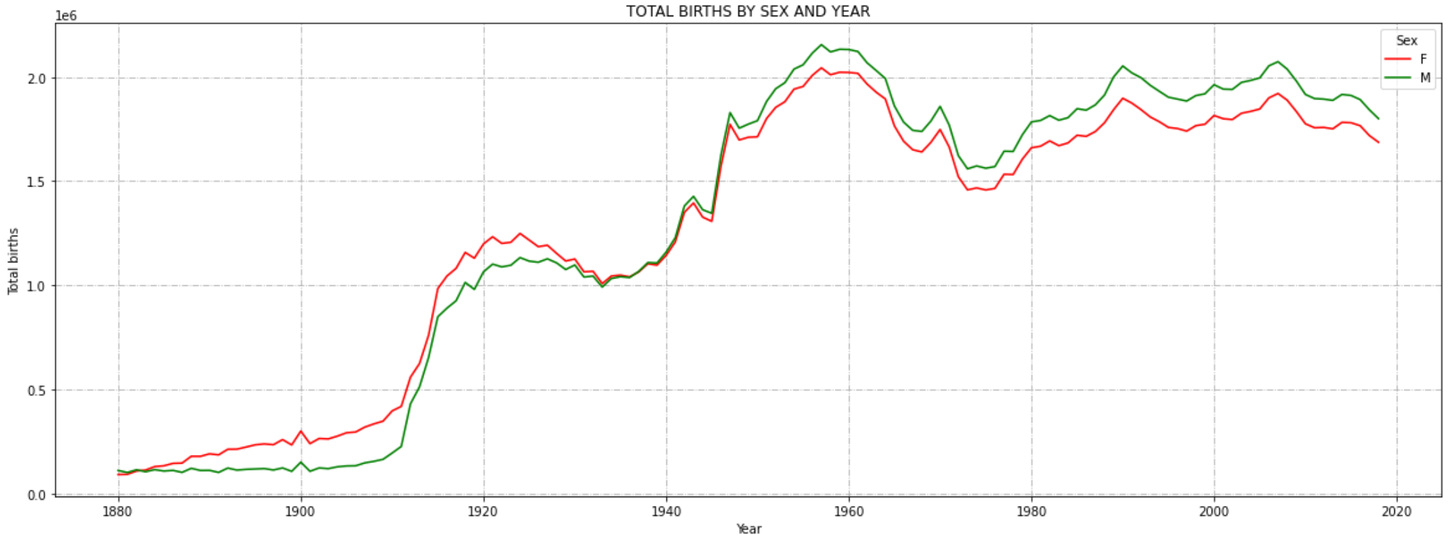

Top 10 tỉnh thành có điểm trung bình môn Tiếng Anh cao nhất
Sở GDĐT TP. Hồ Chí Minh
Sở GDĐT Bình Dương
Sở GDĐT Bà Rịa-Vũng Tàu
Sở GDĐT Hải Phòng
Sở GDĐT Nam Định
Sở GDĐT Hà Nội
Sở GDĐT Đà Nẵng
Sở GDĐT Ninh Bình
Sở GDĐT An Giang
Sở GDĐT Vĩnh Phúc
Tên tỉnh
0,000
2,000
4,000
6,000
8,000
Điểm trung bình


Top 10 tỉnh thành có điểm trung bình môn Tiếng Anh thấp nhất
Sở GDĐT Bắc Kạn
Sở GDĐT Lạng Sơn
Sở GDĐT Trà Vinh
Sở GDĐT Đăk Nông
Sở GDĐT Hậu Giang
Sở GDĐT Cao Bằng
Sở GDĐT Điện Biên
Sở GDĐT Sơn La
Sở GDĐT Hoà Bình
Sở GDĐT Hà Giang
Tên tỉnh
0,000
1,000
2,000
3,000
4,000
5,000
Điểm trung bình

WHY?

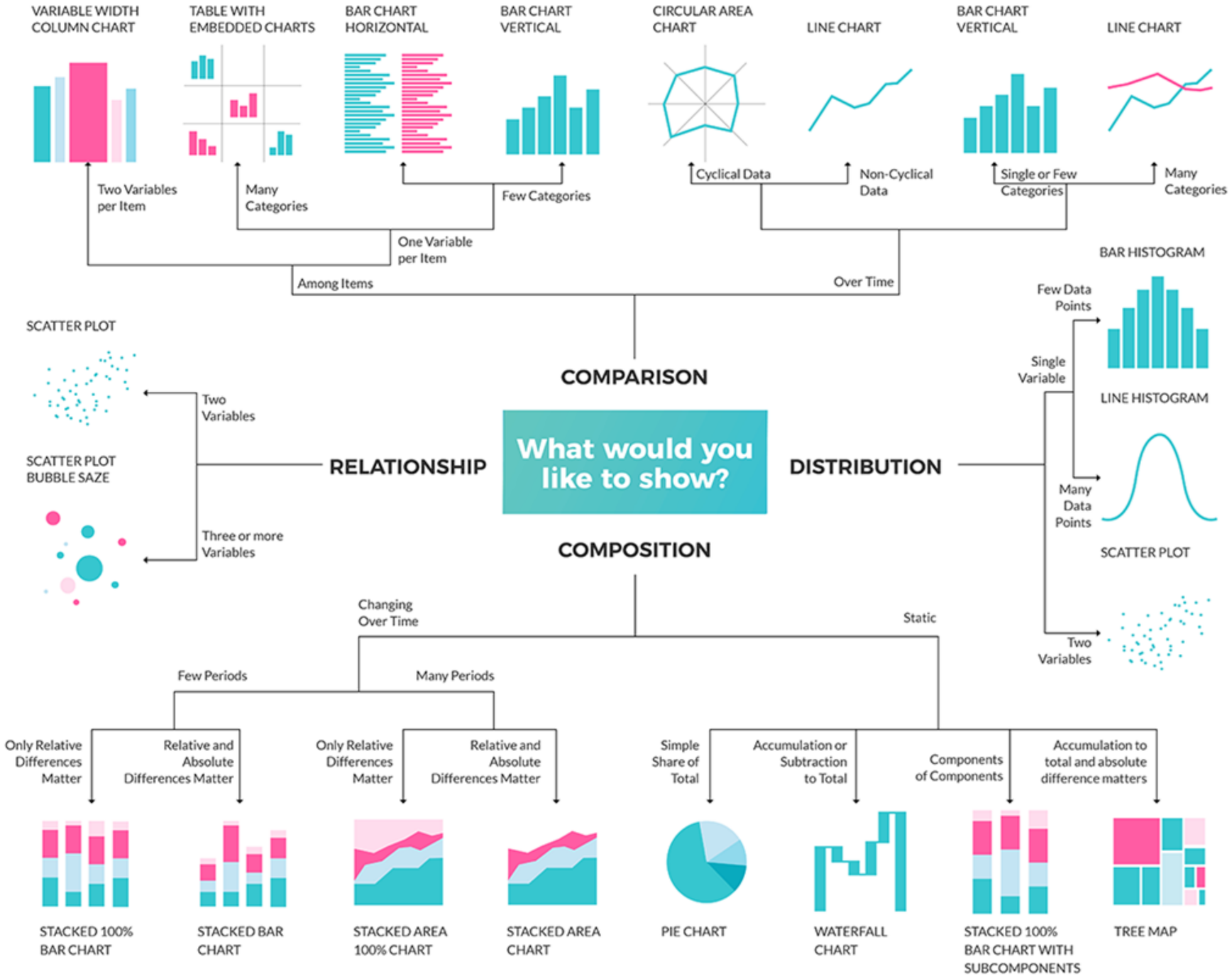
```
df_names.head(5)
```

	Name	Sex	Number	Year
0	Mary	F	7065	1880
1	Anna	F	2604	1880
2	Emma	F	2003	1880
3	Elizabeth	F	1939	1880
4	Minnie	F	1746	1880



Humans are **not** good at **reading numbers**

WHAT?


What would you like to show?
COMPARISON
RELATIONSHIP
DISTRIBUTION
COMPOSITION
VARIABLE WIDTH COLUMN CHART
TABLE WITH EMBEDDED CHARTS
BAR CHART HORIZONTAL
BAR CHART VERTICAL
CIRCULAR AREA CHART
LINE CHART
BAR CHART VERTICAL
LINE CHART
Two Variables per Item
Many Categories
Few Categories
One Variable per Item
Among Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories
Over Time
BAR HISTOGRAM
Few Data Points
Single Variable
LINE HISTOGRAM
Many Data Points
SCATTER PLOT
Two Variables
SCATTER PLOT
Two Variables
SCATTER PLOT BUBBLE SAZE
Three or more Variables
Changing Over Time
Static
Few Periods
Many Periods
Only Relative Differences Matter
Relative and Absolute Differences Matter
Only Relative Differences Matter
Relative and Absolute Differences Matter
Simple Share of Total
Accumulation or Subtraction to Total
Components of Components
Accumulation to total and absolute difference matters
STACKED 100% BAR CHART
STACKED BAR CHART
STACKED AREA 100% CHART
STACKED AREA CHART
PIE CHART
WATERFALL CHART
STACKED 100% BAR CHART WITH SUBCOMPONENTS
TREE MAP

HOW?

How is the score spectrum of English subject in high school graduation exam?

QUESTION

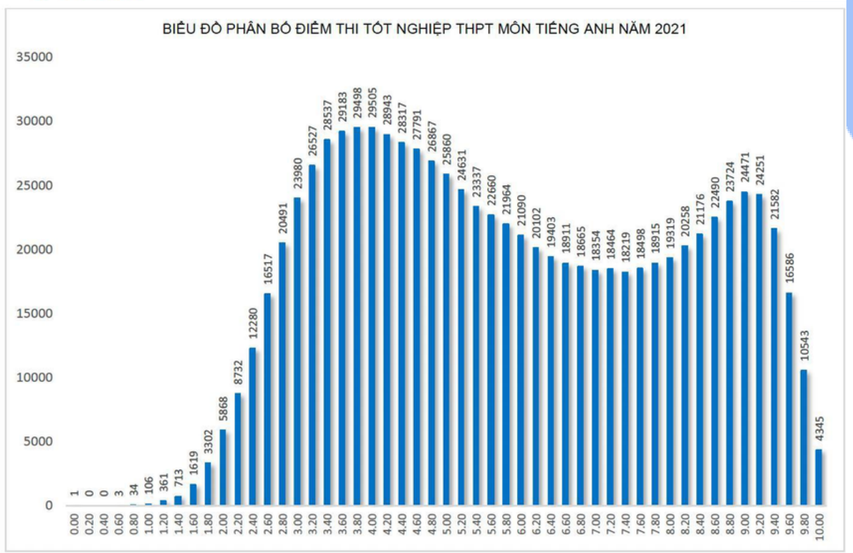
VISUALIZE

ANALYSE

matplotlib

seaborn

pandas



HOW?

QUESTION

Is there any English-level difference between urban and rural areas?

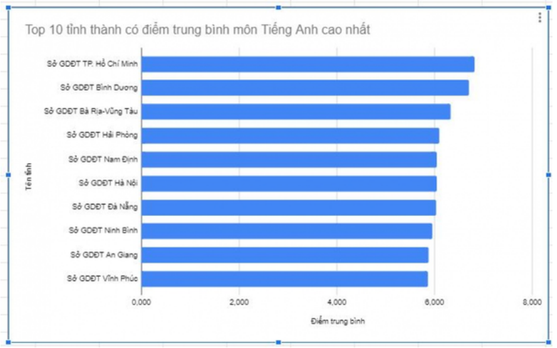
VISUALIZE

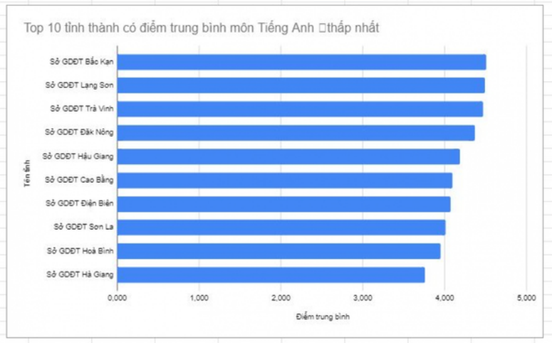
ANALYSE

matplotlib

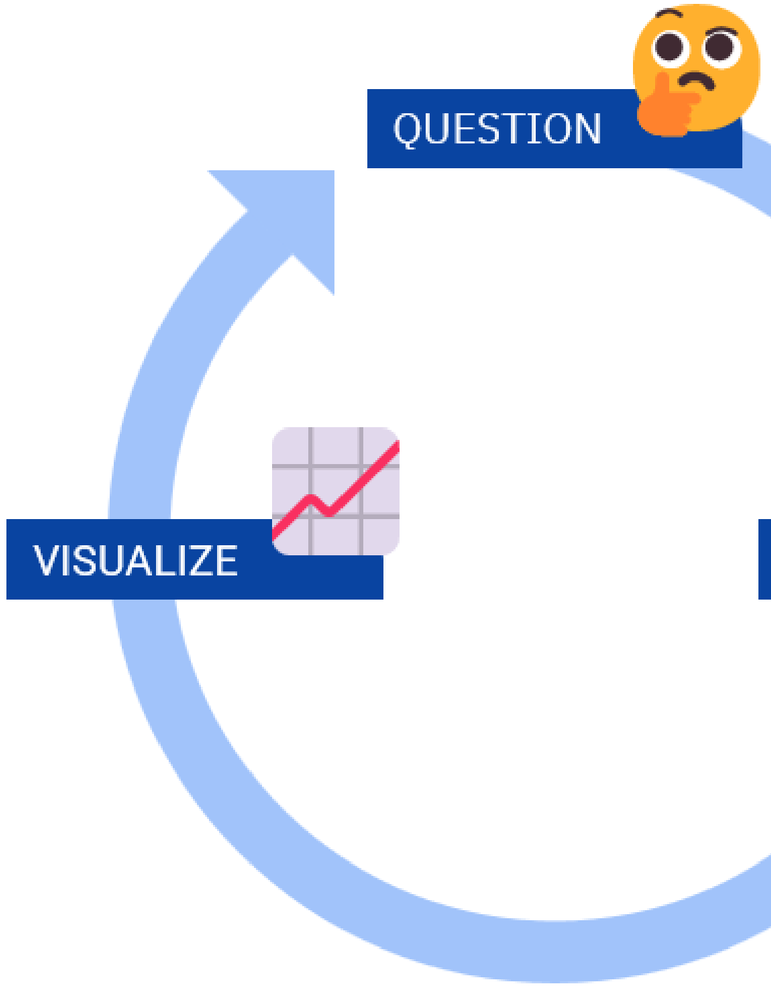
seaborn

pandas





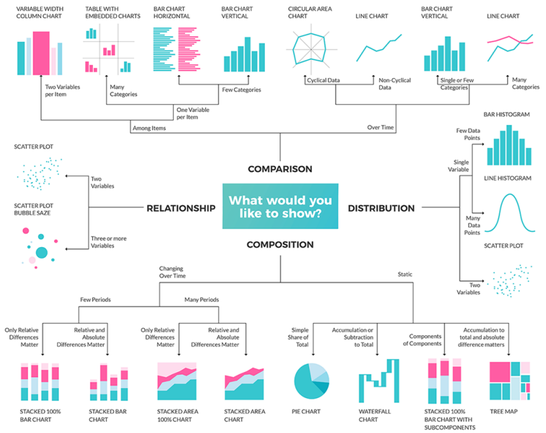
HOW?



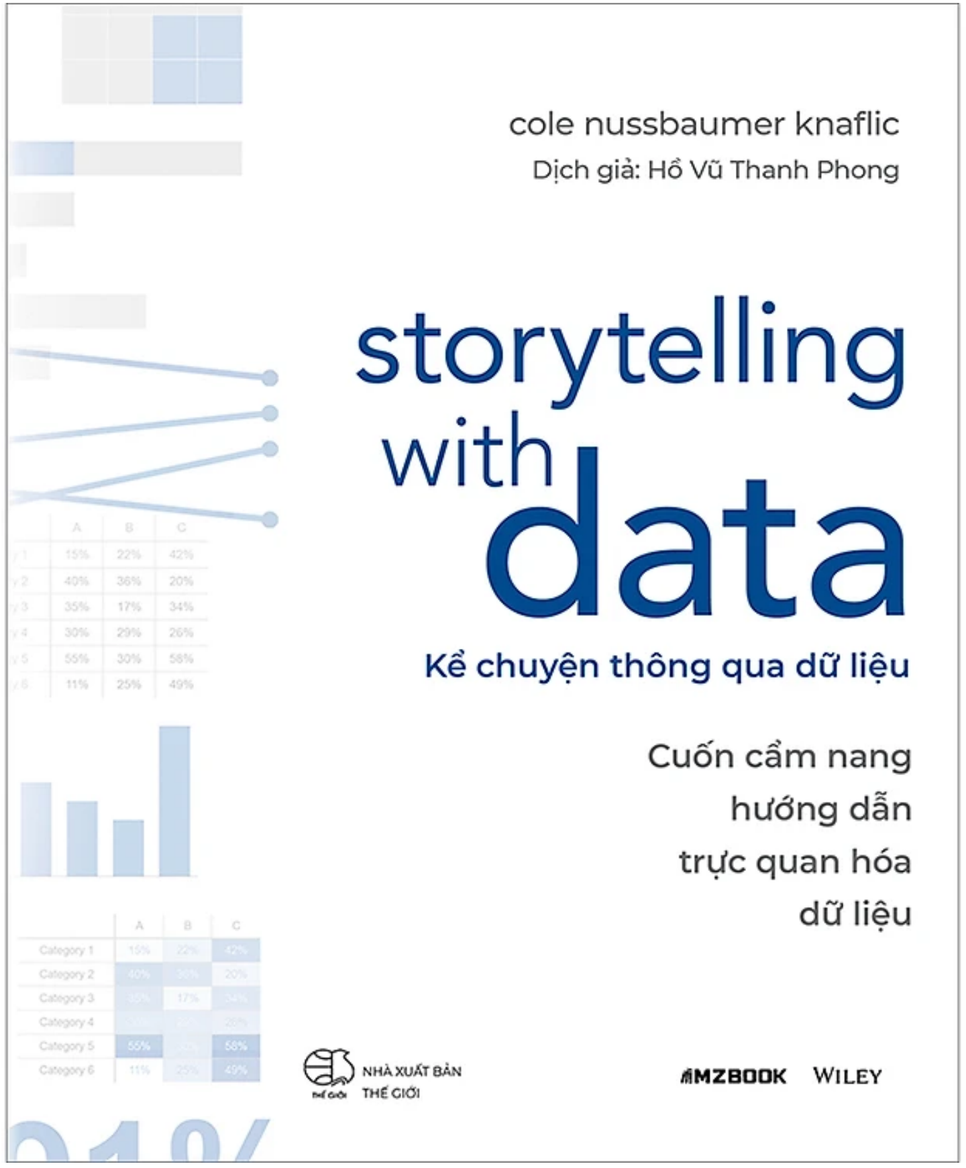
... ? Keep asking until you feel satisfied with the insight gained ☺

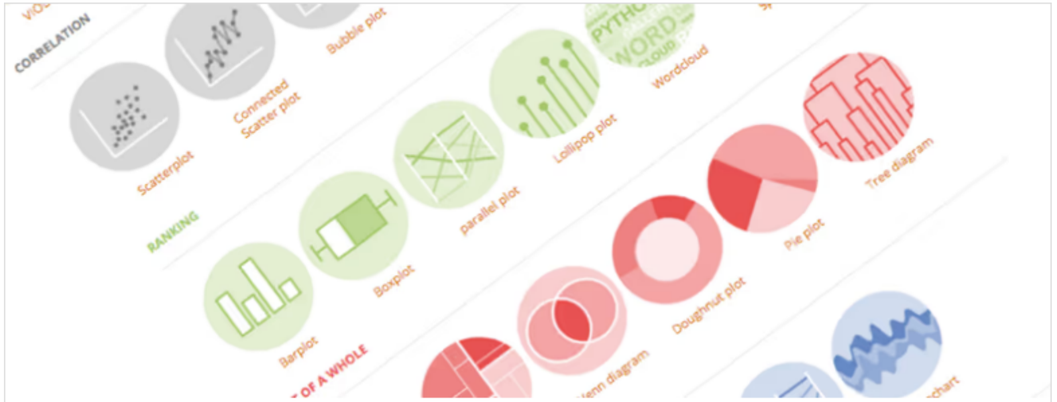
matplotlib

seaborn



pandas





All Charts

A list of more than 300 charts made with Python, coming together with code and explanation

The Python Graph Gallery /