

CITIES AIR QUALITY AND WATER POLLUTION

Instructor: Mr. Nguyễn Việt Anh

Table of Contents

Giới thiệu	1
Thu thập dữ liệu	2
Cách thức	2
Tập dữ liệu	3
Mô tả	4
Phân bố	4
Biểu đồ toàn thế giới về chất lượng nước và không khí	5
Top 10 nước có chất lượng không khí và chất lượng nước kém nhất thế giới	6
Top 10 nước có chất lượng không khí và chất lượng nước tốt nhất thế giới	6
Outliner : Air quality vs Water quality	7
Mối quan hệ giữa avg Air quality và avg Water quality	7
Tình hình tại Việt Nam	8
Những kiến thức được xây dựng từ tập dữ liệu	10
Khoảng tin cậy cho chất lượng nước và không khí của Mỹ	11
Khoảng tin cậy cho chất lượng nước và không khí của Trung Quốc	12
Giả thuyết và suy luận	13
Kết luận	15

Team Members

Họ và tên	Mã số sinh viên
Nguyễn Ngọc Bảo	HE150673
Trần Đại An	HE151259
Đào Minh Đức	HE151256
Cao Văn Cường	HE153674

1. Giới thiệu

- Ô nhiễm không khí và ô nhiễm nước đang đe dọa sức khỏe của người dân ở nhiều nơi trên thế giới.

- Ô nhiễm nguồn nước lan rộng này đang gây nguy hiểm cho sức khỏe của chúng ta. Nước không an toàn giết chết nhiều người hơn mỗi năm so với chiến tranh và tất cả các hình thức bạo lực khác cộng lại

=> Thông qua các báo cáo về mức độ ô nhiễm nước và không khí , nhóm chúng tôi có các câu hỏi sau về tình hình ô nhiễm nước và không khí trên thế giới :

- Tổng quan toàn cầu về mức độ ô nhiễm nước và không khí ?
- Top các quốc gia, thành phố có mức độ ô nhiễm nước và không khí cao nhất ?
- Top các quốc gia, thành phố có mức độ ô nhiễm nước và không khí thấp nhất ?

2. Thu thập dữ liệu

Dữ liệu thu thập vào năm 2020 trên

(<https://www.kaggle.com/cityapiio/world-cities-air-quality-and-water-polution>) từ hơn 200 quốc gia trên thế giới. Từ nguồn đáng tin cậy trên (<https://city-api.io/en>)

Kaggle.com/cityapiio/world-cities-air-quality-and-water-polution

Tìm kiếm

Dữ liệu | Nhiệm vụ | Mã (4) | Thảo luận | Hoạt động | metadata

Tải xuống (236 KB) | Số tài mới

Về tệp này

Dữ liệu ban đầu được lấy từ Nubeo dưới dạng tổng hợp các bình chọn của người dùng.

- Chất lượng không khí thay đổi từ 0 (chất lượng kém) đến 100 (chất lượng tốt hàng đầu)
- Ô nhiễm nước thay đổi từ 0 (không ô nhiễm) đến 100 (ô nhiễm cực độ)

Tập dữ liệu này là một trong những phần công khai của dữ liệu dự án API Thành phố. Cần thêm? Hãy tải toàn bộ dữ liệu của chúng tôi

đ	a	á	à	â	ã	ä	å	æ	ç	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
3796 giá trị duy nhất	"Nước Anh"	11%	"Nước Mỹ"	21%	4%	"Cộng hòa nhân dâ..."	6%	86%	73%	0	100	0	100																			
Thành phố New York	"Newyork"	"Nước Mỹ"	46, 61683773584996	49, 584958495849596																												
Washington DC	"Độc khu Columbia"	"Nước Mỹ"	66, 1298322586451	49, 187142857142854																												
San Francisco	"California"	"Nước Mỹ"	68, 51481869158879	43																												
Berlin	" "	"Nước Đức"	62, 3641384347826	28, 61271676388578																												
Los Angeles	"California"	"Nước Mỹ"	36, 62162162162163	61, 29943582824858																												
Bern	"Canton of Bern"	"Thụy sĩ"	94, 31818181818181	12, 5																												
Geneva	"Bang Geneva"	"Thụy sĩ"	71, 53846153846153	17, 3728813559322																												
Ấn Độ	"Canton of Zürich"	"Thụy Sĩ"	83, 89052309052309	10, 71428571428571																												

Tóm lược

1 tập tin

C .csv

5 cột

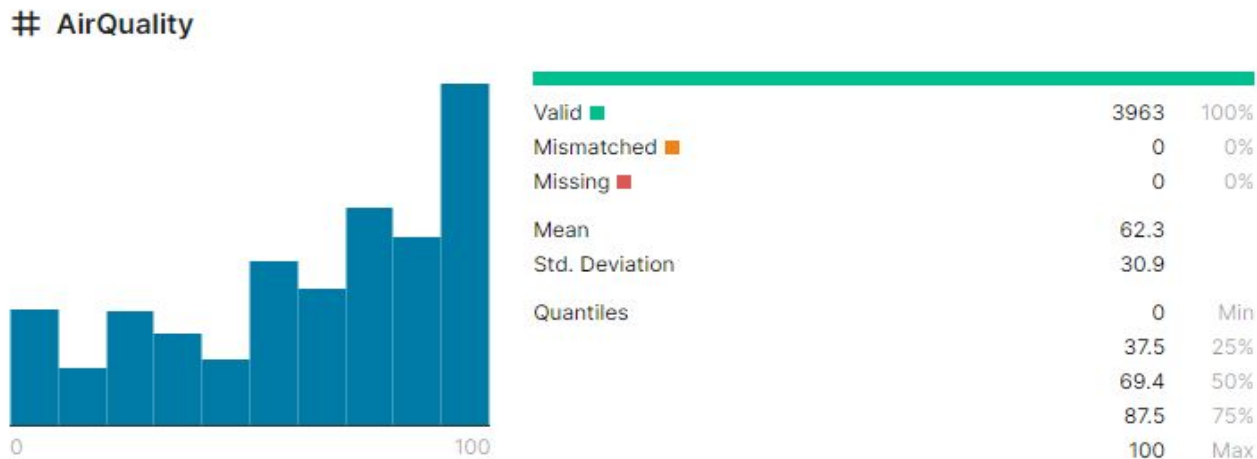
Đầu tiên

3

3. Tập dữ liệu

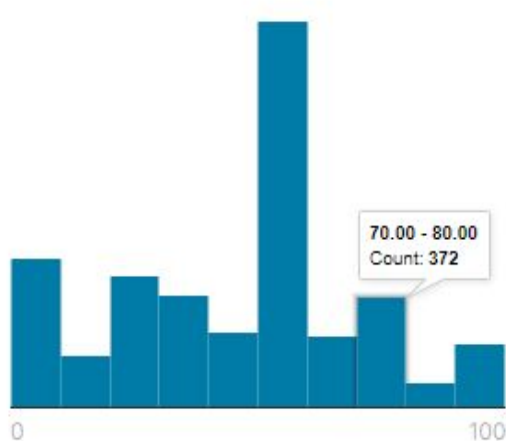
1. Mô tả

- Dữ liệu ban đầu được lấy từ Number dưới dạng tổng hợp các bình chọn của người dùng.
- Min, count, standard deviation, Q1, Q2, Q3 của tất cả các cột số của chất lượng không khí.



- Ô nhiễm nước thay đổi từ 0 (không ô nhiễm) đến 100 (ô nhiễm cực độ).
- Min, count, standard deviation, Q1, Q2, Q3 của tất cả các cột số của chất lượng nước.

WaterPollution

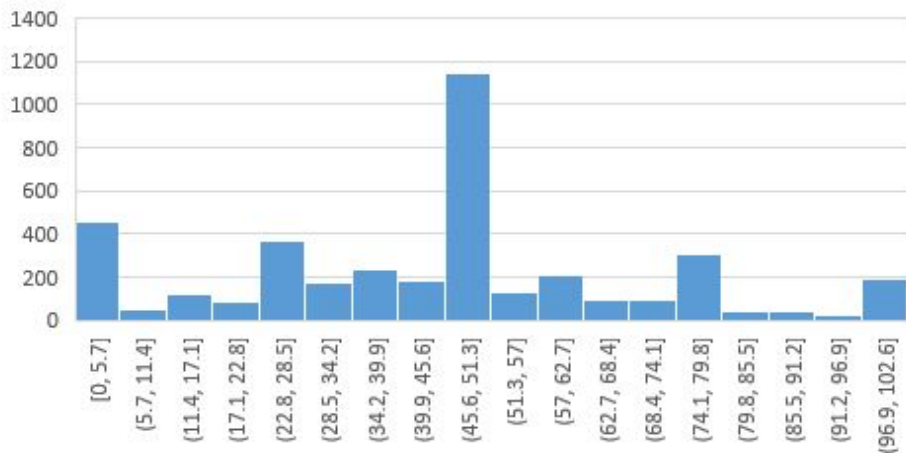


Valid	3963	100%
Mismatched	0	0%
Missing	0	0%
Mean	44.6	
Std. Deviation	25.7	
Quantiles		
	0	Min
	25	25%
	50	50%
	57.7	75%
	100	Max

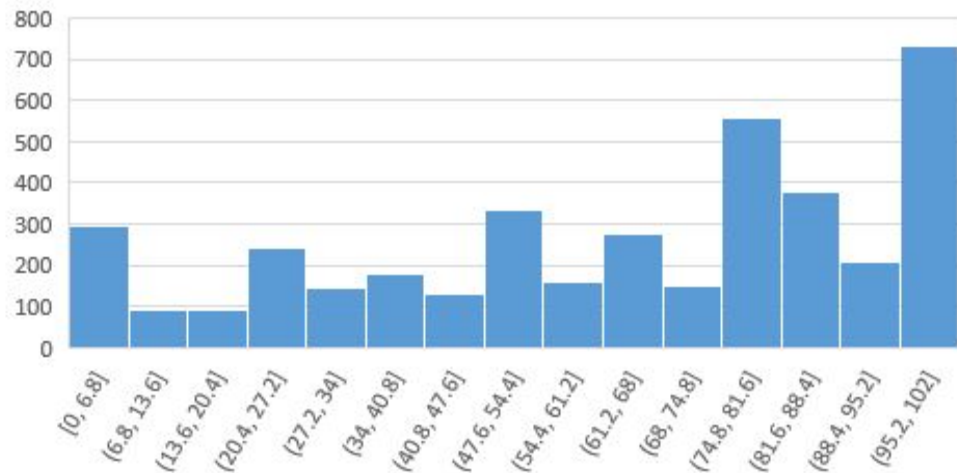
2. Phân bố

- Chất lượng không khí thay đổi từ 0 (chất lượng kém) đến 100 (chất lượng tốt hàng đầu).
- Ô nhiễm nước thay đổi từ 0 (không ô nhiễm) đến 100 (ô nhiễm cực độ).

WaterPollution

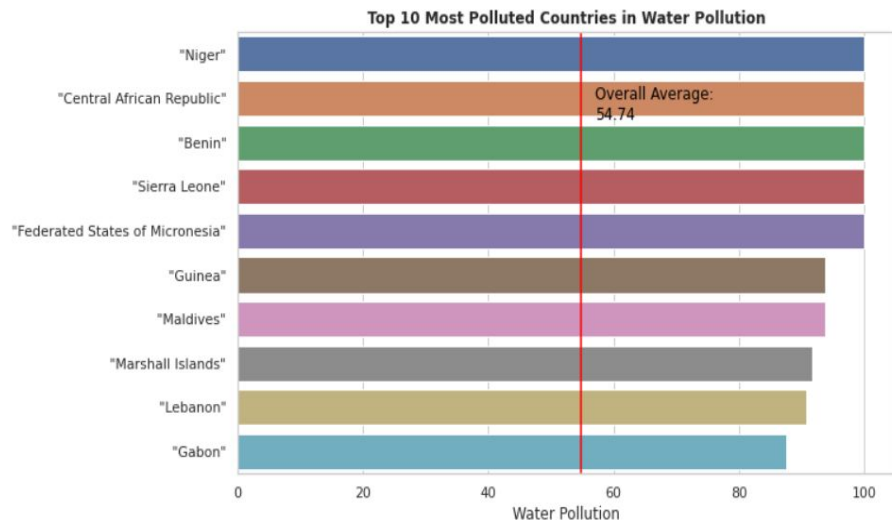
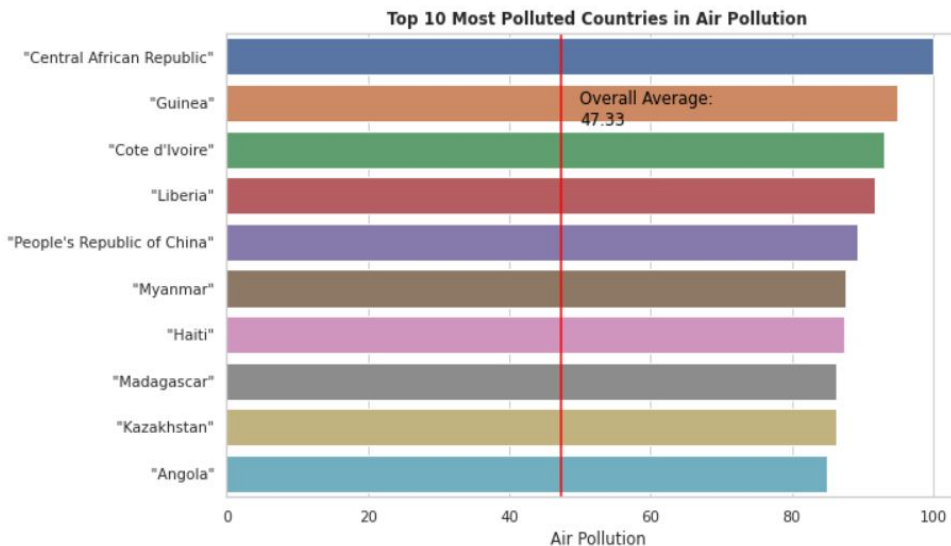


Air Quality



4. Biểu đồ bối cảnh toàn thế giới về chất lượng nước và không khí.

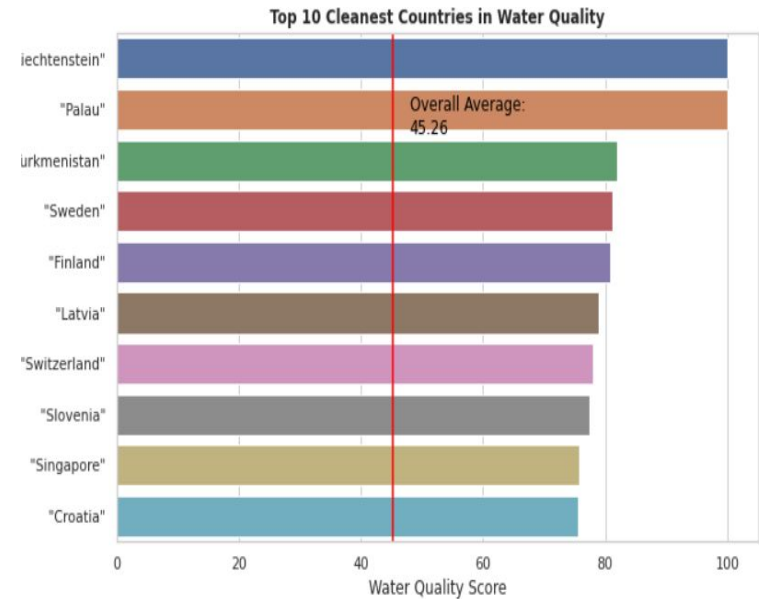
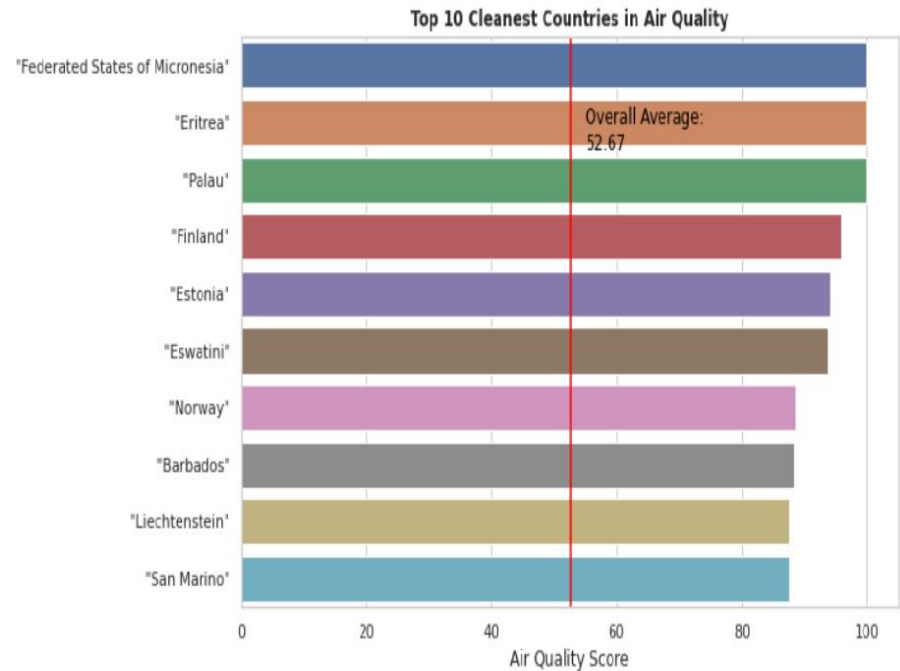
Top 10 nước có chất lượng không khí và chất lượng nước kém nhất thế giới



Nhận xét:

- Cộng hòa Trung Phi là quốc gia ô nhiễm nhất về ô nhiễm không khí.
- Niger, Cộng hòa Trung Phi, Benin, Sierra Leone, và Liên bang Micronesia, được coi là những quốc gia ô nhiễm nhất về ô nhiễm nước.

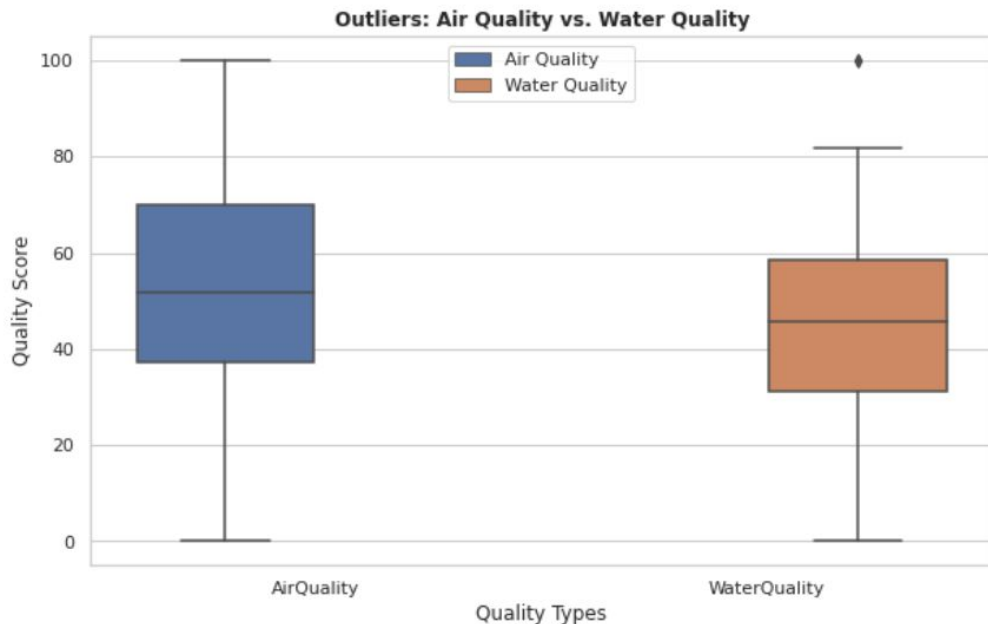
Top 10 nước có chất lượng không khí và chất lượng nước tốt nhất thế giới:



Nhận xét:

- Với chỉ số 100 ba quốc gia Palau, Liên bang Micronesia, Eritrea là các nước được xếp hạng cao nhất trong những nước có chất lượng không khí tốt nhất.
- Palau là quốc gia sạch nhất cả về chất lượng không khí và nước với chỉ số đều được đánh giá là 100 theo dữ liệu trên Kaggle.
- Trái ngược với tình trạng ô nhiễm nguồn nước, Liên bang Micronesia là một trong 3 quốc gia sạch nhất về chất lượng không khí.
- Về chất lượng nước, Liechtenstein và Palau được xếp vào danh sách các quốc gia sạch nhất.

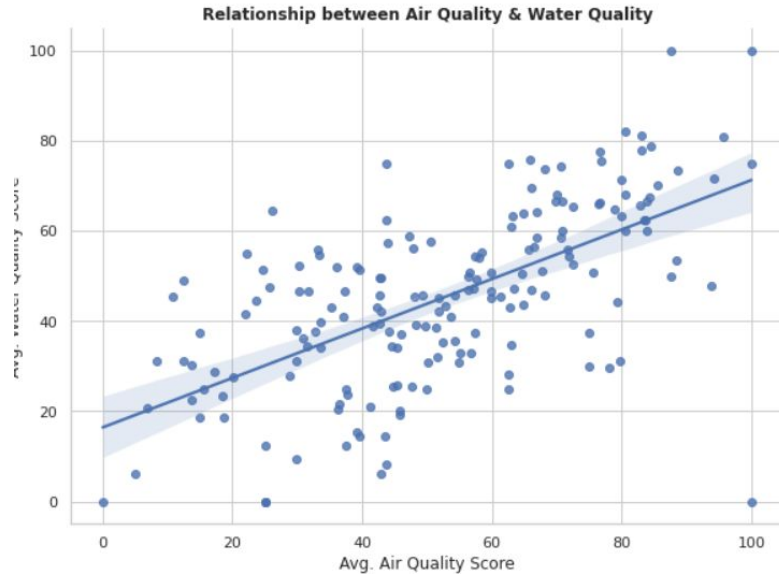
Outliner : Air quality vs Water quality



Nhận xét:

- Chất lượng không khí trung bình gần với normal distribution.
- Chất lượng nước trung bình cũng gần với normal distribution, nhưng có những điểm outlier là Liechtenstein và Palau.

Mối quan hệ giữa avg Air quality và avg Water quality:



Nhận xét:

- Trung bình Chất lượng không khí và Trung bình Chất lượng nước bằng cách nào đó có mối quan hệ tuyến tính và là số dương.
- Sierra Leone, Niger, Cộng hòa Trung Phi và Benin, phải chịu nhiều ô nhiễm khác nhau do tình trạng không ổn định của họ. Ngoài ra, Liên bang Micronesia bị ô nhiễm nguồn nước do phát triển công nghiệp và nông nghiệp - ngược lại, chất lượng không khí tốt.

5. Những kiến thức được xây dựng từ tập dữ liệu.

1. Xây dựng khoảng tin cậy 95% cho chất lượng không khí của nước Mỹ.

- Từ bảng dữ liệu thu thập được , lấy ra dữ liệu về chất lượng không khí của từng thành phố trong nước Mỹ . Để tính toán

khoảng tin cậy 95% cho chất lượng không khí của đất nước Mỹ.

- Với dữ liệu được tính ra : $n = 842$; $\text{mean} = 80.60366$; $\text{standard deviation} = 21.35214$; $\alpha = 0.05$

=> Khoảng dưới và khoảng trên 95% của confidence interval :

- The lower limit of the confidence interval for mean = 79,16
- The upper limit of the confidence interval for mean = 82,05

"AirQuality"	
Mean	80.60366135
Standard Error	0.735843605
Median	87.5
Mode	100
Standard Deviation	21.35214772
Sample Variance	455.9142123
Kurtosis	2.854021961
Skewness	-1.563195601
Range	100
Minimum	0
Maximum	100
Sum	67868.28286
Count	842
Confidence Level(95.0%)	1.444305552

2. Xây dựng khoảng tin cậy 95% cho chất lượng nước của nước Mỹ.

- Từ bảng dữ liệu thu thập được , lấy ra dữ liệu về chất lượng nước của từng thành phố trong nước Mỹ. Để tính toán khoảng tin cậy 95% cho chất lượng nước của Đất nước Mỹ.

Với dữ liệu được tính ra : $n = 842$; $\text{mean} = 40.05239$; $\text{standard deviation} = 24.8797$; $\alpha = 0.05$

=> Khoảng dưới và khoảng trên 95% của confidence interval :

- The lower limit of the confidence interval for mean = 38,37
- The upper limit of the confidence interval for mean = 41,73

"WaterPollution"	
Mean	40.0523965
Standard Error	0.85741317
Median	42.9029304
Mode	50
Standard Deviation	24.8797604
Sample Variance	619.002477
Kurtosis	-0.2279119
Skewness	0.17505702
Range	100
Minimum	0
Maximum	100
Sum	33724.1179
Count	842
Confidence Level(95.0%)	1.68292092

3. Xây dựng khoảng tin cậy 95% cho chất lượng không khí của nước Trung Quốc.

- Từ bảng dữ liệu thu thập được, lấy ra dữ liệu về chất lượng không khí của từng thành phố trong nước Trung Quốc Để tính toán khoảng tin cậy 95% cho chất lượng nước của Đất nước Trung Quốc.

Với dữ liệu được tính ra : $n = 238$; $\text{mean} = 10.7610$; $\text{standard deviation} = 15.8932$; $\alpha = 0.05$

=> Khoảng dưới và khoảng trên 95% của confidence interval :

- The lower limit of the confidence interval for mean = 8.7315.
- The upper limit of the confidence interval for mean = 12.7905.

"AirQuality"	
Mean	10.76106703
Standard Error	1.030208286
Median	0
Mode	0
Standard Deviation	15.89327936
Sample Variance	252.5963288
Kurtosis	2.885826561
Skewness	1.70940915
Range	75
Minimum	0
Maximum	75
Sum	2561.133952
Count	238
Confidence Level(95.0%)	2.029535059

4. Xây dựng khoảng tin cậy 95% cho chất lượng nước của nước Trung Quốc.

- Từ bảng dữ liệu thu thập được, lấy ra dữ liệu về chất lượng nước của từng thành phố trong nước Trung Quốc để tính toán khoảng tin cậy 95% cho chất lượng không khí của Đất nước Trung Quốc.

Với dữ liệu được tính ra : $n = 238$; $\text{mean} = 54.6174$;
 $\text{standard deviation} = 15.7803$; $\alpha = 0.05$

=> Khoảng dưới và khoảng trên 95% của confidence interval :

- The lower limit of the confidence interval for mean = 52.6023.
- The upper limit of the confidence interval for mean = 56.6325.

"WaterPollution"	
Mean	54.6174883
Standard Error	1.02288922
Median	50
Mode	50
Standard Deviation	15.7803663
Sample Variance	249.01996
Kurtosis	3.78722233
Skewness	0.70078748
Range	100
Minimum	0
Maximum	100
Sum	12998.9622
Count	238
Confidence Level(95.0%)	2.01511632

6. Giả thuyết và suy luận:

1. Giả thuyết 1:

- Mỹ và Trung Quốc là hai cường quốc có GDP hàng đầu thế giới

TRUNG QUỐC

Xếp hạng GDP

1st (danh nghĩa; 2021)

2nd (PPP; 2021)

MỸ

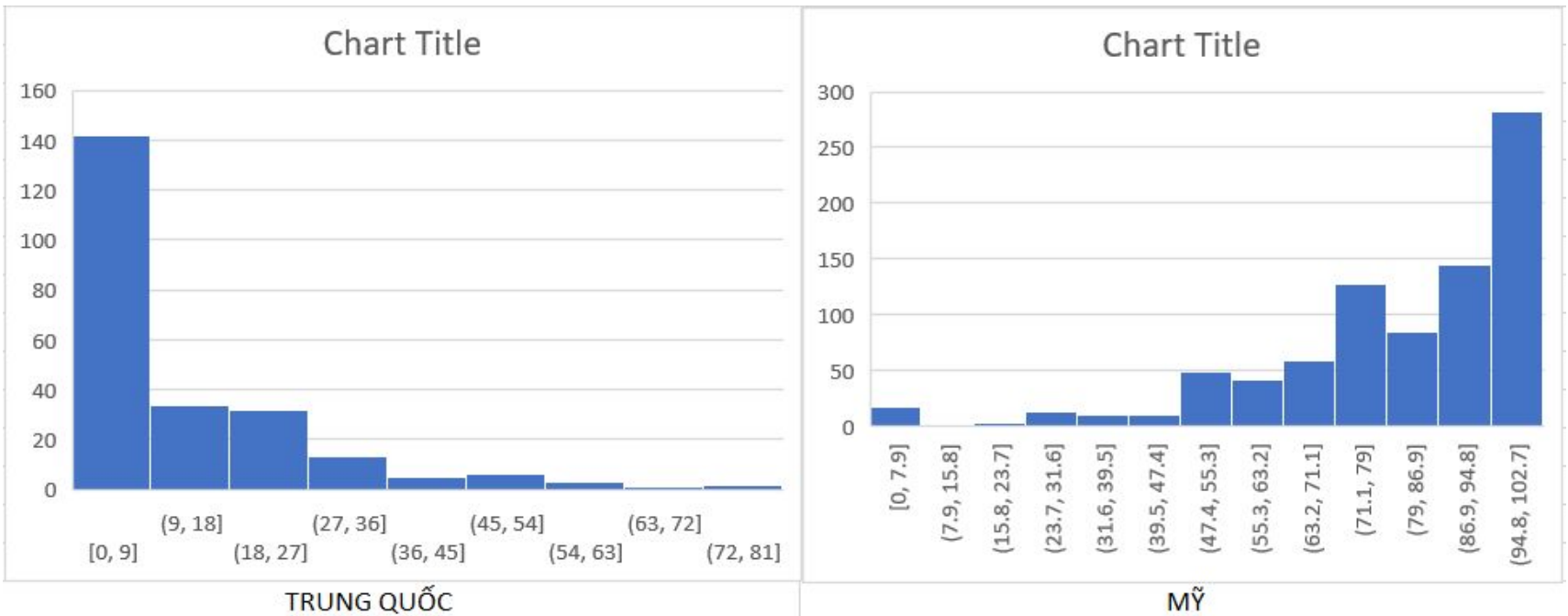
Xếp hạng GDP

2nd (danh nghĩa; 2021)

1st (PPP; 2021)

Nhưng chất lượng không khí tại Mỹ có thể được đánh giá cao hơn Trung Quốc rất nhiều.

- Dữ liệu thu nhập được:



- Từ dữ liệu thu thập được ta có:

$$n_1 = 842, \sigma_1 = 21.352$$

$$n_2 = 238, \sigma_2 = 15.893$$

$$\bar{x}_1 - \bar{x}_2 - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

=>Lower confidence limit = 67.361

=>Upper confidence limit = 72.322

-Từ kết quả trên với khoảng tin cậy 95% về sự khác biệt giữa chỉ số chất lượng không khí trung bình của 2 nước Mỹ và Trung Quốc. Ta thấy chất lượng không khí trung bình tại Mỹ hơn chất lượng không khí trung bình tại Trung Quốc từ 67.361 đến 72.322 chỉ số

=> Suy luận: Từ nhận định trên, chúng ta có thể dễ dàng nhận thấy rằng mặc dù là 2 nước có nền kinh tế rất giàu mạnh nhất nhì thế giới nhưng chất lượng không khí tại các thành phố thì quá khác biệt, chỉ số chất lượng không khí trung bình tại Mỹ tốt hơn rất nhiều so với Trung Quốc.

2. Giả thuyết 2:

-Mặc dù chất lượng không khí trung bình tại Mỹ và Trung Quốc chênh lệch nhau rất lớn, nhưng chúng ta muốn biết liệu chất lượng không khí trung bình tại top 3 thành phố lớn nhất của Trung Quốc có hơn chất lượng không khí trung bình tại top 3 thành phố lớn nhất của Mỹ không. Biết chỉ số chất lượng không khí trung bình top 3 thành phố lớn nhất của Trung Quốc là $x_1 = 20.904$, còn tại Mỹ là $x_2 = 46.9567$. Sử dụng $\alpha = 0.05$

1. Chỉ số chất lượng không khí khác nhau là $\mu_1 - \mu_2$

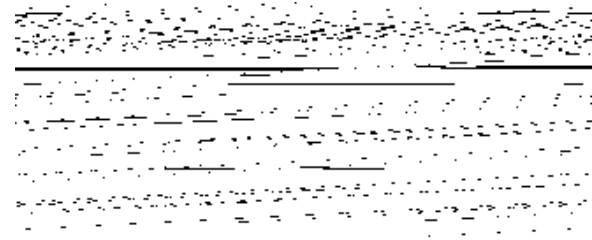
2. $H_0: \mu_1 - \mu_2 = 0$ hoặc $\mu_1 = \mu_2$ và $\Delta_0 = 0$

3. $H_1: \mu_1 < \mu_2$. Chúng ta sẽ bác bỏ H_0 nếu chất lượng không khí trung bình tại top 3 thành phố lớn nhất của Trung Quốc có hơn chất lượng không khí trung bình tại top 3 thành phố lớn nhất của Mỹ

5. The test statistic is:

$$\Rightarrow z = -4.001$$

\Rightarrow Bác bỏ H_0 : vì $-4.001 < 1.645 \Leftrightarrow Z_0 < Z_{\alpha/2}$



* Suy luận: Ta thấy lý do chỉ số chất lượng không khí tại Trung Quốc kém hơn rất nhiều so với Mỹ không chỉ đến từ những thành phố có nhiều khu công nghiệp sản xuất mà chính những thành phố lớn, thành phố trọng điểm, chất lượng không khí cũng chênh lệch rất lớn với Mỹ. Chính phủ Trung Quốc cần có biện pháp để cải thiện tình trạng đáng báo động này.

7.Kết luận - Những kiến thức thu được từ dự án.

- Cách thu thập dữ liệu một cách hiệu quả , đáng tin cậy nhất.
- Cách sử dụng Excel và các công cụ khác để phân tích và giải quyết các bài toán xác suất với tập dữ liệu lớn.
- Hiểu thêm về những kiến thức mới.
- Kỹ năng giao tiếp và làm việc nhóm.