Internet Prices around 200+ countries in 2022.

(1 GB Internet Prices (from cheapest to expensive) across the world.)

INTRODUCTION

A commercial use of the Internet was introduced in the 1980's, but the world did not know that this would change the course of human history. The Internet has opened a new world for many people around the world and it has become a vital part of human life. All the tiring offline jobs are upgraded to online.Since the origin of the internet to today's 5G internet, there have been significant changes in its usage, pricing, and end user base.

This document contains the 5 phases of analysis. Ask Phase, it's the first phase of the data analysis process. It helps clarify the purpose of the analysis as well as the kind of problem we're trying to solve. Following the Prepare phase, we collect the data relevant to the problem we are trying to solve.Third Phase of analysis is Process. The data cleaning process is an important step in this phase, as well as finding the errors, inaccuracies, and inconsistencies in the data.The last phases are analyze and share phase, in analyze phase the primary goal is to find the relationships, trends, and patterns from the dataset. Creating the visualization, getting findings from the visualization and sharing the insights with stakeholders are included in the share phase.

All the phases are done with two programming languages and libraries like MySQL, Python and Matplot library.

Ask Phase

The major question that need to be solved in this analysis are:

- The price variation in the years 2020 and 2021.
- Compare the count of internet users in the years 2020 and 2021.
- Trend in Internet price, speed, end users in both countries and continents across the globe.

Prepare Phase

The updated dataset used in this analysis is from kaggle, which has thousands of open dataset (link).The dataset (Internet price across the world) used in this analysis has Internet price(for 1gb), speed, cheapest and most expensive price, number of users and population for each country and continent. This phase is done in MySQL language. For importing data into MySQL, create a new table with the total number of columns equal and in the same order to the dataset that is downloaded from kaggle.

###OPTIONAL##----------

```
-- data_analysis.internet_price definition

CREATE TABLE `internet_price` (
  `SlNo` int DEFAULT NULL,
  `CountryCode` varchar(50) DEFAULT NULL,
  `Country` varchar(50) DEFAULT NULL,
  `Continent` varchar(50) DEFAULT NULL,
  `no_of_plans` double DEFAULT NULL,
  `avg_price_1gb` double DEFAULT '0',
  `cheapest_price` double DEFAULT NULL,
  `expensive_price` double DEFAULT NULL,
  `avg_price_2021` double DEFAULT NULL,
  `avg_price_2020` double DEFAULT NULL,
  `internet_users` int DEFAULT NULL,
  `population` int DEFAULT NULL,
  `user_percentage` double DEFAULT NULL,
  `avg_speed` double DEFAULT NULL
)
```

—----------

	SlNo	CountryCode	Country	Continent	no of plans	avg price 1gb	cheapest price
1	0	IL	Israel	NEAR EAST	27	0.05	0.02
2	1	KG	Kyrgyzstan	CIS (FORMER USSR)	20	0.15	0.1
3	2	FJ	Fiji	OCEANIA	18	0.19	0.05
4	3	IT	Italy	WESTERN EUROPE	29	0.27	0.09
5	4	SD	Sudan	SUB-SAHARAN AFRICA	33	0.27	0.03
6	5	RU	Russia	CIS (FORMER USSR)	22	0.29	0.13
7	6	MD	Moldova	EASTERN EUROPE	18	0.32	0.07
8	7	BD	Bangladesh	ASIA (EX. NEAR EAST)	60	0.34	0.11
9	8	LK	Sri Lanka	ASIA (EX. NEAR EAST)	60	0.38	0
10	9	CL	Chile	SOUTH AMERICA	59	0.39	0.24
11	10	FR	France	WESTERN EUROPE	45	0.41	0.09
12	11	ID	Indonesia	ASIA (EX. NEAR EAST)	53	0.42	0.17
13	12	BY	Belarus	CIS (FORMER USSR)	14	0.43	0.02
14	13	SM	San Marino	WESTERN EUROPE	2	0.43	0.24
15	14	VN	Vietnam	ASIA (EX. NEAR EAST)	46	0.49	0.07
16	15	DZ	Algeria	NORTHERN AFRICA	20	0.51	0.16
17	16	CN	China	ASIA (EX. NEAR EAST)	36	0.52	0.3
18	17	WS	Samoa	OCEANIA	42	0.56	0.19
19	18	PK	Pakistan	ASIA (EX. NEAR EAST)	60	0.59	0.06
20	19	KZ	Kazakhstan	CIS (FORMER USSR)	19	0.59	0.25

expensive price	avg price 2021	avg price 2020	internet users	population	user percentage	avg speed
20.95	0.11	0.9	6,788,737	8,381,516	80.997	28.01
7.08	0.21	0.27	2,309,235	6,304,030	36.631	16.3
0.85	0.59	3.57	452,479	883,483	51.215	25.99
3.54	0.43	1.73	50,540,000	60,627,291	83.362	37.15
0.92	0.63	0.68	12,512,639	41,801,533	29.933	9.5
1.86	0.52	0.91	124,000,000	145,734,038	85.087	20.46
2.79	1.12	2.82	3,083,783	4,051,944	76.106	29.46
2.22	0.7	0.99	129,180,000	166,303,498	77.677	10.43
5.53	0.51	0.78	7,121,116	21,228,763	33.545	13.15
1.83	0.71	1.87	14,864,456	18,729,160	79.365	22.49
118.2	0.81	1.21	59,470,000	64,990,511	91.506	60.94
2.94	0.64	2.99	196,000,000	267,670,543	73.224	17.7
16.62	0.89	2.36	7,048,231	9,452,617	74.564	9.99
0.63	1.16	6.86	20,100	33,785	59.494	[NULL]
5.55	0.57	1.31	68,172,134	95,545,962	71.35	33.9
2.24	0.65	5.15	26,350,000	42,228,408	62.399	12.44
1.21	0.61	9.89	1,010,740,000	1,427,647,786	70.798	83.43
6.86	10.86	30.09	66,023	196,129	33.663	[NULL]
8.59	0.69	1.85	118,800,000	213,756,286	55.577	16.73
2.35	0.46	0.49	13,913,699	18,319,618	75.95	18.71

Initially, data types for a few columns were not appropriate for analysis, as some of the columns contained non integer values. There are some columns which need to be deleted and added in the dataset

Process Phase

This phase drives the analysis, all the insights that need to be found mostly depend on this phase. Data cleaning is done in this phase, fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

Eliminating the duplicate values from the dataset

Delete the duplicate value to improve the accuracy of finding the analysis.

#Finding the duplicate values and deleting it from dataset

```
SELECT *
FROM internet_price ip z
group by CountryCode having count(*) > 1;
```

S.N	Country cod	Country	Continental region	NO. OF Internet Plar	Average price of 1GB (USD)	Cheapest 1GB for 30 days (USD	Mo
174	AW	Aruba (Netherlands)	CARIBBEAN	17	4.44	0.74	
175	AW	Aruba (Netherlands)	CARIBBEAN	17	4.44	0.74	
182	LB	Lebanon	NEAR EAST	15	4.81	1.21	
183	LB	Lebanon	NEAR EAST	15	4.81	1.21	

#SQL code for deleting duplicate values.

```
DELETE
FROM internet_price
WHERE SlNo = 175 or SlNo =231 or SlNo =183;
```

Adding columns to the dataset.

A column can be added to a dataset if it enhances the analysis and helps gain insight..In this dataset we have data related to number of users and population for each country, if the percentage of internet users are found based on population, countries with highest internet users can be found.

Calculating the percentage of users in each country with a given population and Internet user and the values are added to a new column called ‘user_percentage’.

#Adding a column user_percentage into the table

```
ALTER TABLE internet_price
ADD user_percentage DOUBLE AFTER population;
```

Checking the possible errors that can occur during calculation will eliminate the error percentage.

#Checking the possible error in user and population

```
SELECT internet_users,population
FROM internet_price ip
WHERE internet_users > population ;
```

#Updating values to the column user_percentage

```
UPDATE internet_price
SET user_percentage = ((internet_users/population)*100)
WHERE population !=0;
```

Changing the error values/non-integer values.

There are columns which have non-integer values,which need to be changed first and only then values can be converted to its specific data type and then calculation is possible. To attain that non-integer values need to be replaced by 0 or NULL

#Updating the values of the column ‘avg_price_1gb’ to 0 if any row contains non-integer values.

```
UPDATE internet_price SET avg_price_1gb = 0
where avg_price_1gb = "NO PROVIDERS" OR avg_price_1gb =
"HYPERINFLATION" OR avg_price_1gb = "Prices listed in
non-convertible 'units'" OR
avg_price_1gb = "UNRELIABLE EXCHANGE RATES" ;
```

#Updating the values of the column ‘ internet price’ which some of the rows contains “NULL” value, by converting those values to 0 calculation in the future of the analysis will be easy.

```
UPDATE internet_price set avg_price_2020 = 0
where (avg_price_2020 = " ”)
```

Once all the values are corrected, it's time to convert the data types of the columns to its appropriate types.

#Changing the data type of avg_price_2021, avg_price_2020, ‘avg_price_1gb’ to Double

```
ALTER TABLE internet_price
MODIFY avg_price_2021 DOUBLE DEFAULT 0;

ALTER TABLE internet_price
MODIFY avg_price_2020 DOUBLE DEFAULT 0;

ALTER TABLE internet_price
MODIFY avg_price_1gb DOUBLE DEFAULT 0;
```

Analyze Phase

After extensive data cleaning, it's time to analyze the data to get insights.Using this trusted data, analysis is done in the language Python. To access the code in python, here is the. link