Worldwide Internet Users Analysis

Data analyst Project

Jithin Kurian

07/10/2022

INTRODUCTION

Internet usage is one of the most required services in the 21st century. The world without internet can be very challenging. This project is to identify key information regarding internet usage around the globe. The data is about internet usage, cost and speed across 200+ countries. And the data is collected from various sources in the internet. The project is focusing on answering some questions outlined. The project gives some inputs to global internet providers, analysts, experts in internet services etc…

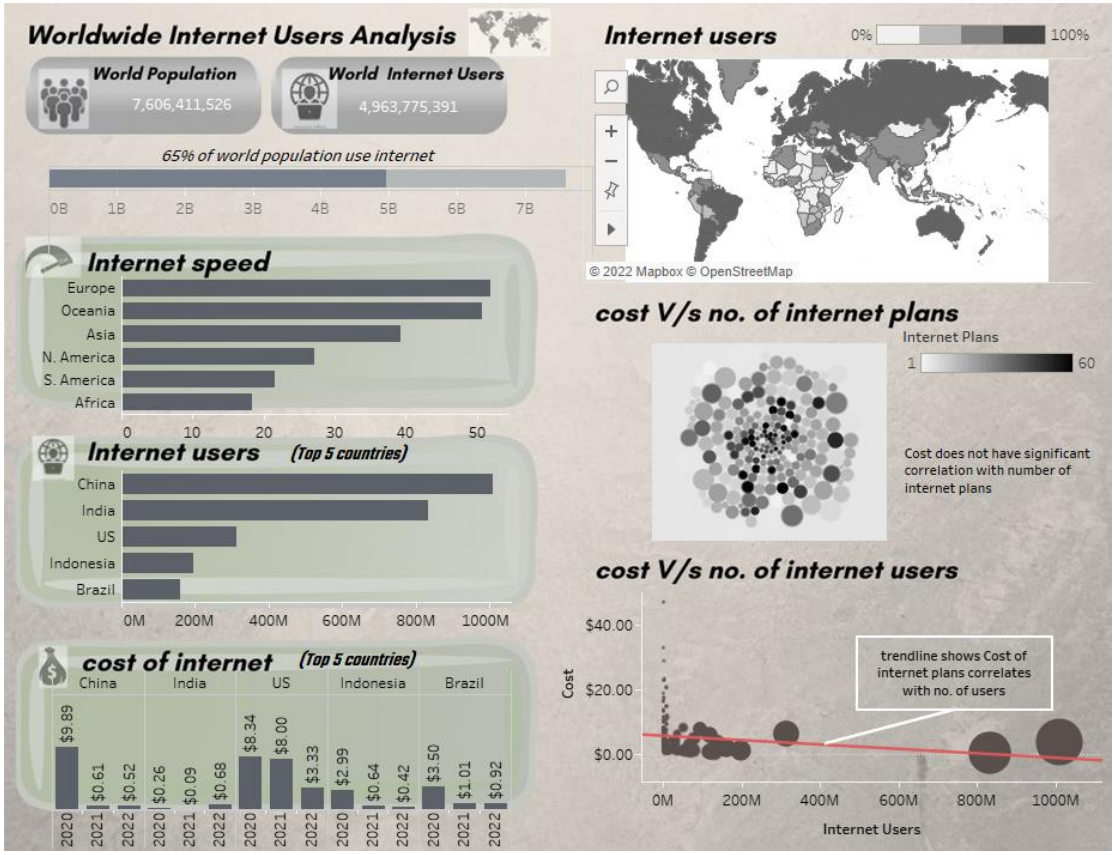
EXECUTIVE SUMMARY

The given screenshot of Tableau dashboard is the final analysis result.

Link to Tableau dashboard :-

(https://public.tableau.com/views/worldwideinternetusersanalysis/worldwideinternetusersanalysis?:language=en-US&:display_count=n&:origin=viz_share_link

)



QUESTIONS

- What is the percentage of world population use internet?
- Top 5 countries in terms of number of internet users
- What percentage of population use internet categorised by continent?
- Which continent with highest share of internet users?
- Does number of internet plans correlate with cost of plans?
- Is there any huge difference in average internet speed in continents?
- Does every country in a continent have same internet speed? Or varies heavily?
- How the cost of internet plans in 2020, 2021, 2022 over countries? (let's take top countries as sample)
- Does the cost of plans correlates with number of internet users?
- Are there any countries with less than 50 % of population use internet?

HYPOTHESIS

My assumptions on the analysis are as follows:-

- Around 75% of world population use internet
- I assume US, China, Australia, India and Germany are the highest in rank when counting number of internet users
- I assume over 70% of population use internet over continent wiseThe data is about internet usage
- Asia would be having highest share of internet users
- I think number of internet plans highly correlates with cost of plans
- I assumes Europe and North America have high internet speed and Asia and Africa will be lowest with significant difference
- My hypothesis on speed among countries in same continent would be around same value
- There wouldn't be any difference in cost of internet plans among the top countries during these years

- There would be significant relationship between cost of internet and number of internet users
- I assume there is no country with internet users which is less than 50% of population

DATA CLEANING

Before the analysis I have done the cleaning the data. Below I have provided screenshots of the description and steps in data cleaning.

```
--data cleaning
--table 1
--Avoiding NULL value found in population
SELECT *
FROM internet_users
WHERE population IS NOT Null


--checking any miss leading value
--(easy way to find out wrong info on population and internet users count ie. more users than population)

SELECT country_name, region, internet_users, population
,round ((internet_users/population :: numeric)*100,2) percentage_of_population_use_internet
FROM internet_users
WHERE population IS NOT Null
ORDER BY percentage_of_population_use_internet DESC
```

```
--correcting region into continent
--(step 1)

SELECT sub_region, region, COUNT(sub_region)
FROM internet_users
WHERE population IS NOT Null
GROUP BY sub_region, region

--(step 2)
SELECT country_name, sub_region,region, internet_users, population
,round ((internet_users/population :: numeric)*100,2) percentage_of_population_use_internet,
CASE when sub_region = 'Northern America' then 'North America'
     when sub_region = 'Caribbean' then 'North America'
     when sub_region = 'Central America' then 'North America'
     when sub_region = 'South America' then sub_region
     else region
END continent
FROM internet_users
WHERE population IS NOT Null
```

```sql
-- extracting required information from table

SELECT country_name,
CASE when sub_region = 'Northern America' then 'North America'
     when sub_region = 'Caribbean' then 'North America'
     when sub_region = 'Central America' then 'North America'
     when sub_region = 'South America' then sub_region
     else region
END continent, internet_users, population
,round ((internet_users/population :: numeric)*100,2) percentage_of_population_use_internet
FROM internet_users
WHERE population IS NOT Null

--creating new cleaned table
DROP TABLE IF EXISTS internet_users_cleaned

CREATE TABLE internet_users_cleaned
(
    country_name  VARCHAR(255) PRIMARY KEY,
    continent VARCHAR(255),
    internet_users INT,
    population INT,
    percentage_of_population_use_internet INT
)

INSERT INTO internet_users_cleaned
( SELECT country_name,
CASE when sub_region = 'Northern America' then 'North America'
     when sub_region = 'Caribbean' then 'North America'
     when sub_region = 'Central America' then 'North America'
     when sub_region = 'South America' then sub_region
     else region
END continent, internet_users, population
,round ((internet_users/population :: numeric)*100,2) percentage_of_population_use_internet
FROM internet_users
WHERE population IS NOT Null
)
```

```sql
--cleaning table 2
--finding and avoiding NULL values

SELECT *
FROM internet_speed
WHERE avg_internet_speed is NULL

DELETE from internet_speed
WHERE avg_internet_speed is NULL
```

```sql
--cleaning table 3

SELECT *
FROM internet_prices
ORDER BY internet_plans

DELETE
FROM internet_prices
WHERE internet_plans <=0

SELECT *
FROM internet_prices
ORDER BY Average_price_of_1GB_2022

SELECT *
FROM internet_prices
WHERE Average_price_of_1GB_2021 is null

SELECT *
FROM internet_prices
WHERE Average_price_of_1GB_2020 is null

UPDATE internet_prices
SET Average_price_of_1GB_2020 = 19
WHERE country_name = 'Bermuda'
```

```sql
--avoiding outlier
SELECT*
FROM internet_prices
ORDER BY internet_prices.most_expensive_1gb DESC

DELETE
FROM internet_prices
WHERE internet_prices.most_expensive_1gb >750

--alter column names more readable
ALTER TABLE internet_prices RENAME COLUMN average_price_of_1gb_2022 TO avg_cost_2022;
ALTER TABLE internet_prices RENAME COLUMN average_price_of_1gb_2021 TO avg_cost_2021
ALTER TABLE internet_prices RENAME COLUMN average_price_of_1gb_2020 TO avg_cost_2020
```

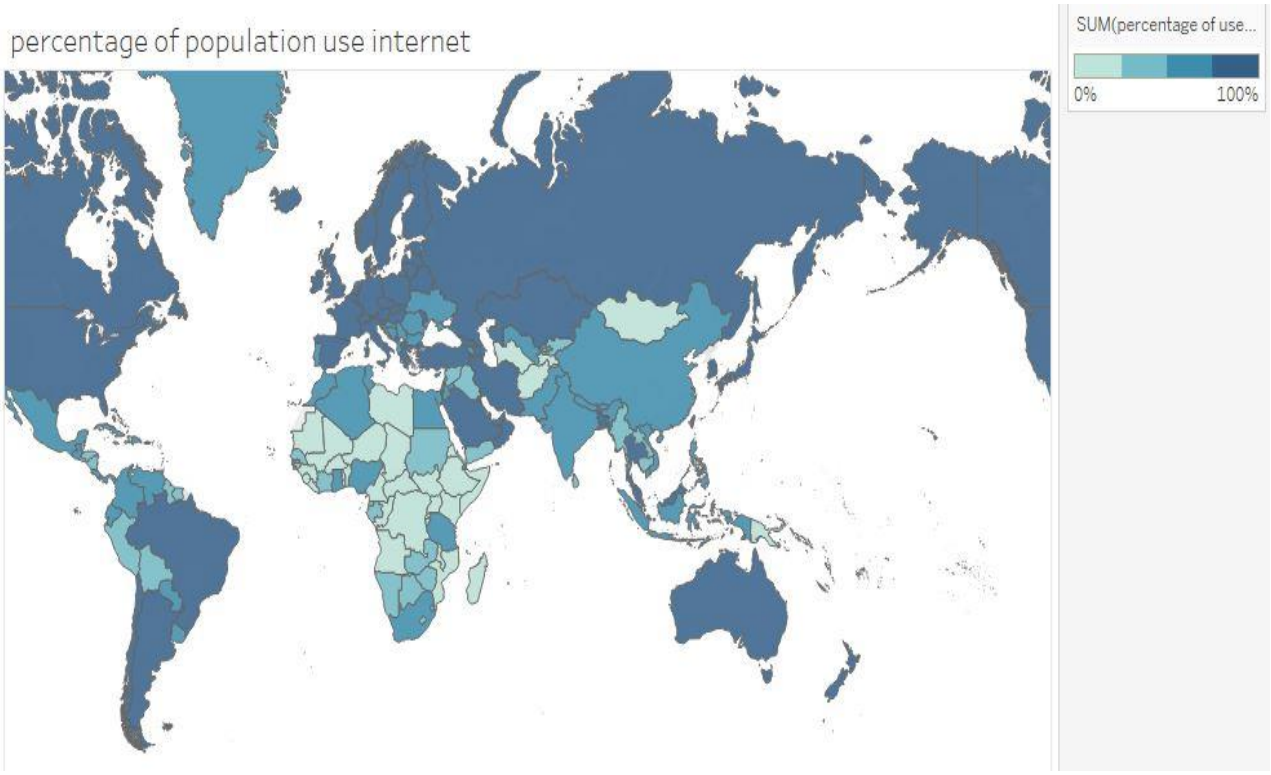
```sql
-- getting needed info
SELECT country_name, internet_plans, avg_cost_2022, avg_cost_2021, avg_cost_2020
,round((avg_cost_2022 + avg_cost_2021 + avg_cost_2020)/3,2) as avg_cost_of_all_years
,cheapest_1gb_for_30_days as cheapest_plan, most_expensive_1gb as most_expensive_plan
FROM internet_prices
```

```sql
--making a cleaned table
CREATE TABLE internet_prices_cleaned
(
    country_name  VARCHAR(255) PRIMARY KEY,
    internet_plans INT,
    avg_cost_2022 DECIMAL,
    avg_cost_2021 DECIMAL,
    avg_cost_2020 DECIMAL,
    avg_cost_of_all_years DECIMAL,
    cheapest_plan DECIMAL,
    most_expensive_plan DECIMAL
)

INSERT INTO internet_prices_cleaned
(
    SELECT country_name, internet_plans, avg_cost_2022, avg_cost_2021, avg_cost_2020
    ,round((avg_cost_2022 + avg_cost_2021 + avg_cost_2020)/3,2) as avg_cost_of_all_years
    ,cheapest_1gb_for_30_days as cheapest_plan, most_expensive_1gb as most_expensive_plan
    FROM internet_prices
)
```

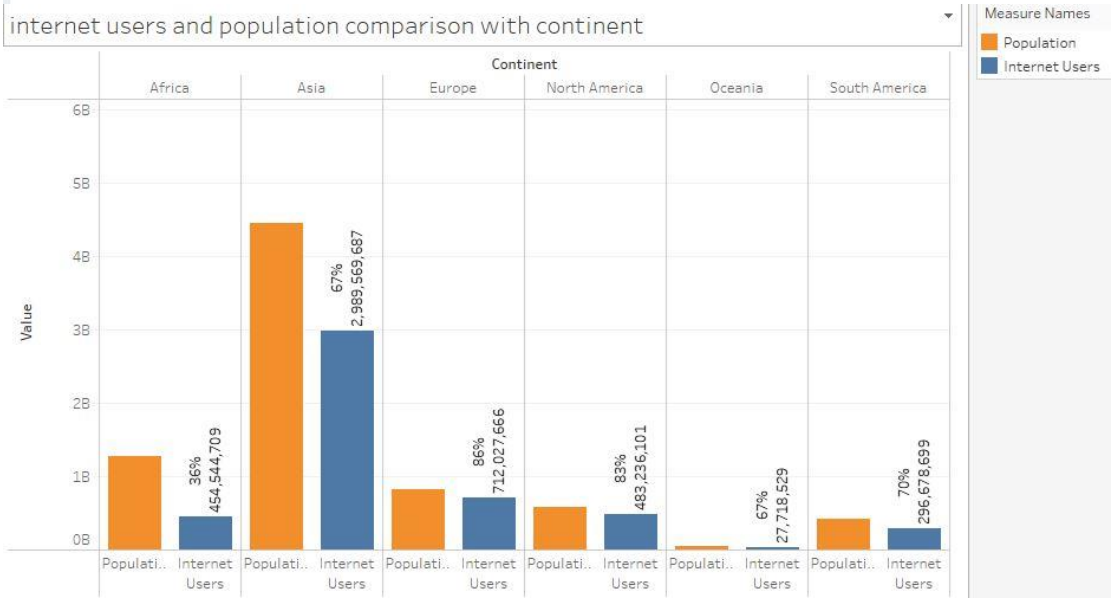
DATA ANALYSIS

After the data cleaning, I have conducted analysis of the data. Below I have provided the screenshots of SQL queries for analysis and screenshots of the visuals of the analysis from Tableau.

```sql
-- percentage of polulation use internet
SELECT country_name, internet_users, percentage_of_population_use_internet
FROM internet_users_cleaned
ORDER BY internet_users DESC
```



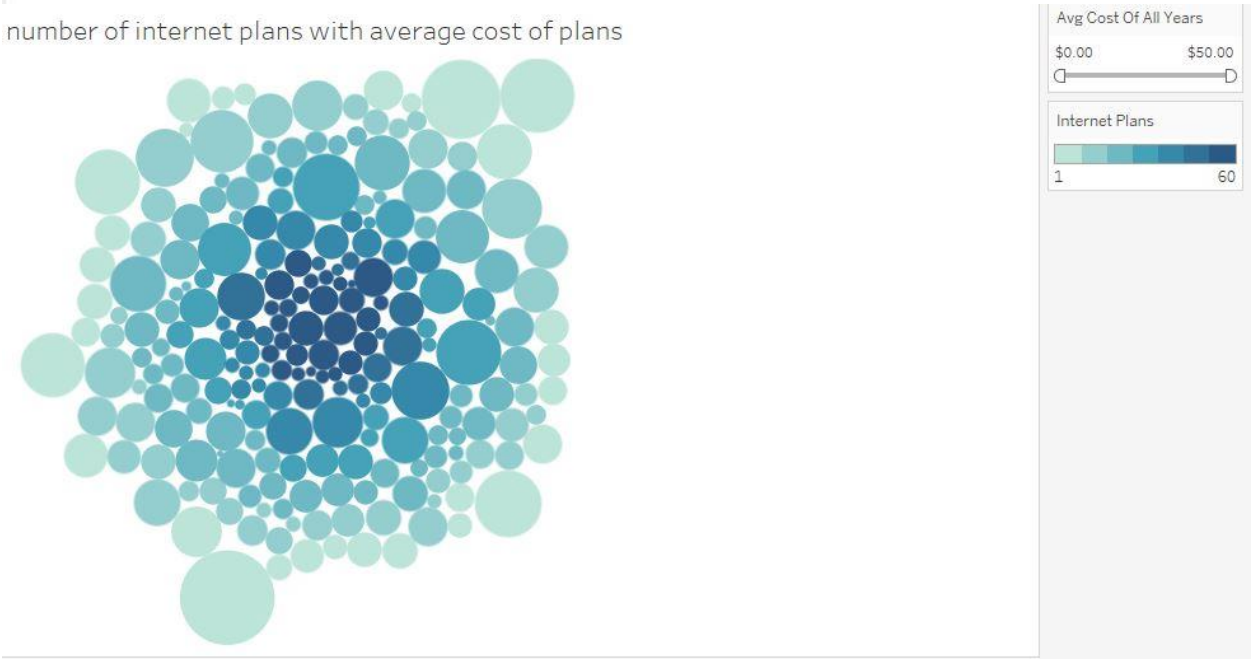
```sql
-- percentage of population use internet continent wise
SELECT DISTINCT continent, SUM (population) OVER (PARTITION by continent) as population_continent
,SUM(internet_users) OVER (PARTITION BY continent) as internet_users_continent,
round (((SUM(internet_users) OVER (PARTITION BY continent)/ SUM(population)
        OVER (PARTITION by continent) :: Decimal)) * 100,0) as percent_of_population_use_internet
FROM internet_users_cleaned

--or by using window
WITH users_continentCTE AS
(
    SELECT DISTINCT continent, SUM (population) OVER (PARTITION by continent) as population_continent
    ,SUM(internet_users) OVER (PARTITION BY continent) as internet_users_continent
    FROM internet_users_cleaned
)

SELECT *,ROUND((internet_users_continent/population_continent:: DECIMAL)*100,0) AS percent_of_population_use_internet
FROM users_continentCTE
```




```sql
-- second model
SELECT DISTINCT continent, SUM (internet_users)
FROM internet_users_cleaned
GROUP BY continent
```

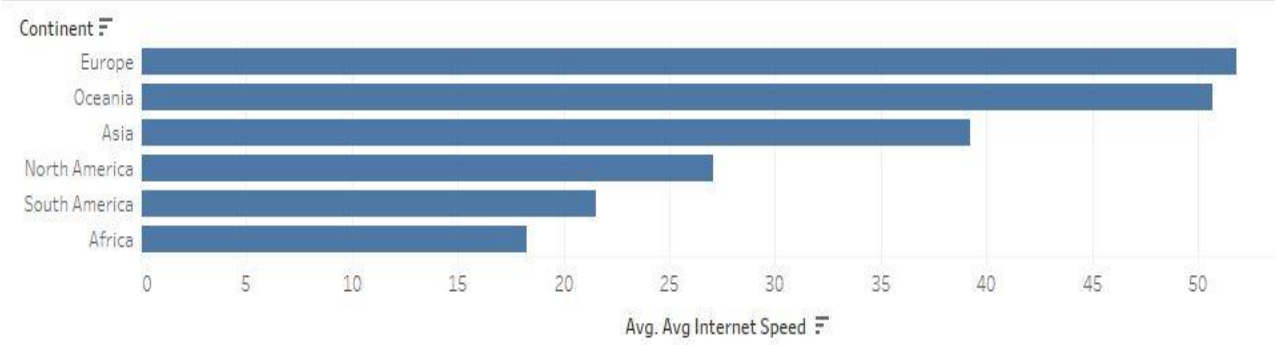
```sql
--number of plans with average cost of plan
SELECT country_name,internet_plans, avg_cost_of_all_years
FROM internet_prices_cleaned
ORDER BY internet_plans desc, avg_cost_of_all_years
```

number of internet plans with average cost of plans



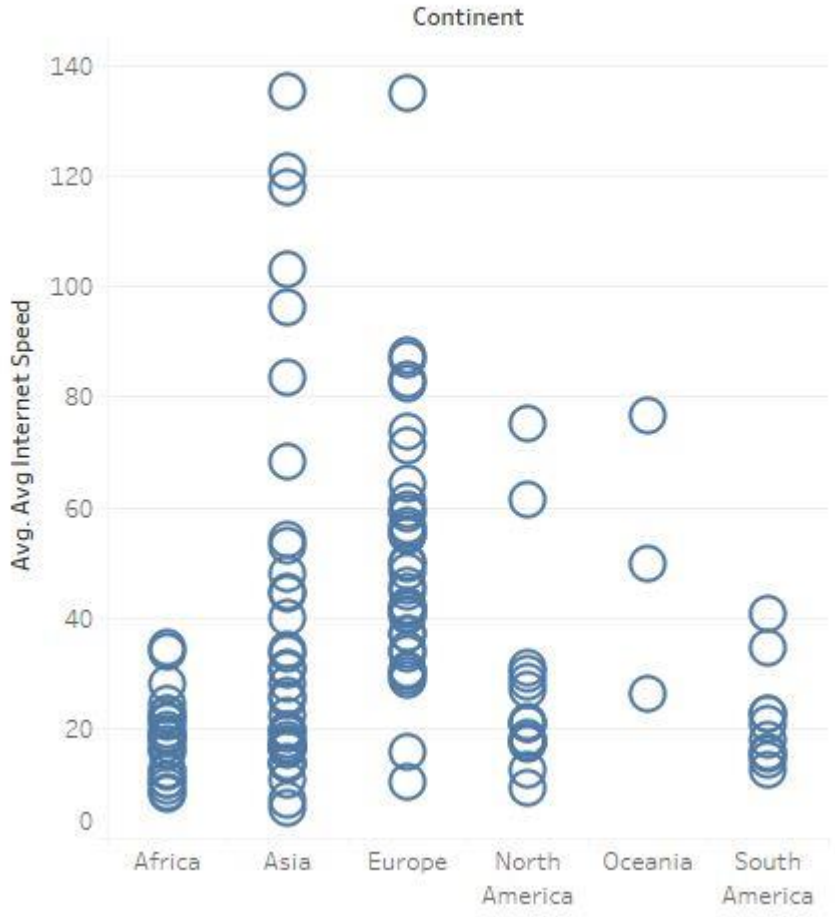

```sql
--continent wise internet speed
SELECT DISTINCT us.continent, round (AVG (sp.avg_internet_speed),0) as avg_internet_speed
FROM internet_users_cleaned us
JOIN internet_speed sp
ON us.country_name = sp.country_name
GROUP BY us.continent
ORDER BY avg_internet_speed DESC
```

internet speed with continent



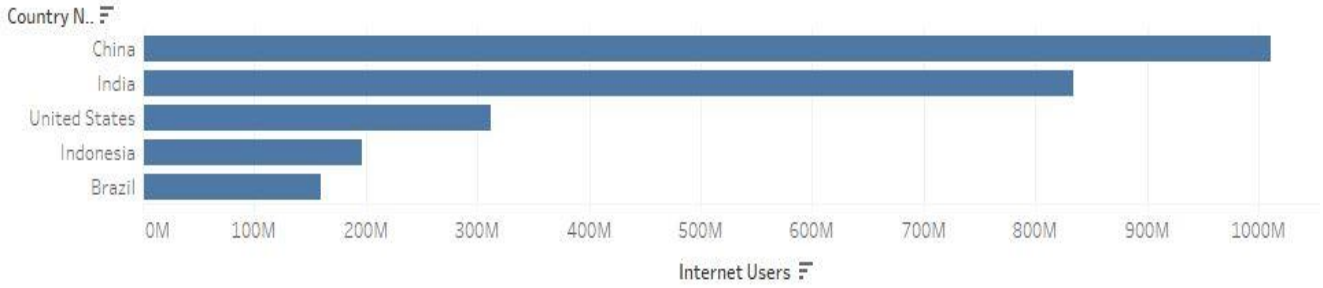
```sql
--finding countries speed difference in continent
SELECT DISTINCT us.continent, us.country_name, sp.avg_internet_speed
FROM internet_users_cleaned us
JOIN internet_speed sp
ON us.country_name = sp.country_name
GROUP BY us.continent,us.country_name, sp.avg_internet_speed
ORDER BY  us.continent
```

internet speed difference in continent



```sql
--top 5 highest internet users countries
SELECT country_name, internet_users
FROM internet_users_cleaned
ORDER BY internet_users DESC
LIMIT 5
```

top 5 countries with number of internet users

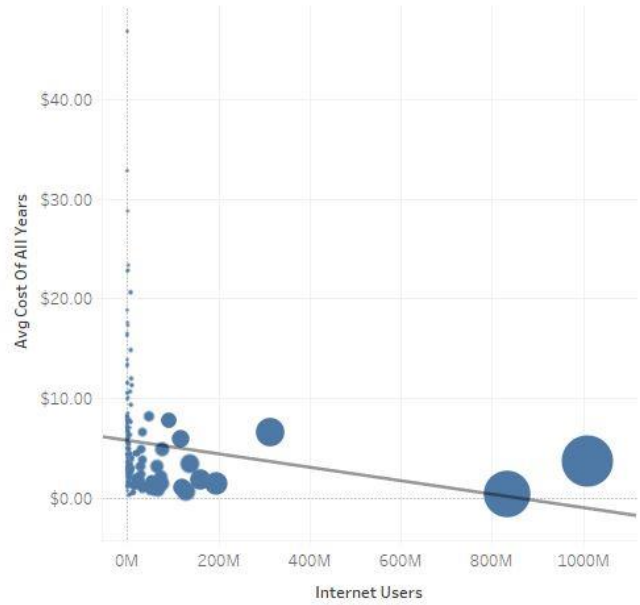


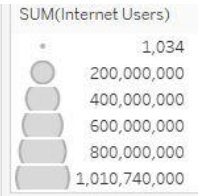

```sql
-- top 5 highest internet users countries' internet cost in 2020, 2021 and 2022
SELECT us.country_name, ps.avg_cost_2020, ps.avg_cost_2021, ps.avg_cost_2022
FROM internet_users_cleaned us
JOIN internet_prices_cleaned ps
ON ps.country_name = us.country_name
ORDER BY us.internet_users DESC
LIMIT 5
```

Top 5 countries with variation in cost of plan in years

```sql
-- finding correlation between number of internet users and average cost of internet plans by countries
SELECT us.country_name, ps.avg_cost_of_all_years
FROM internet_users_cleaned us
JOIN internet_prices_cleaned ps
ON ps.country_name = us.country_name
ORDER BY us.internet_users DESC
```

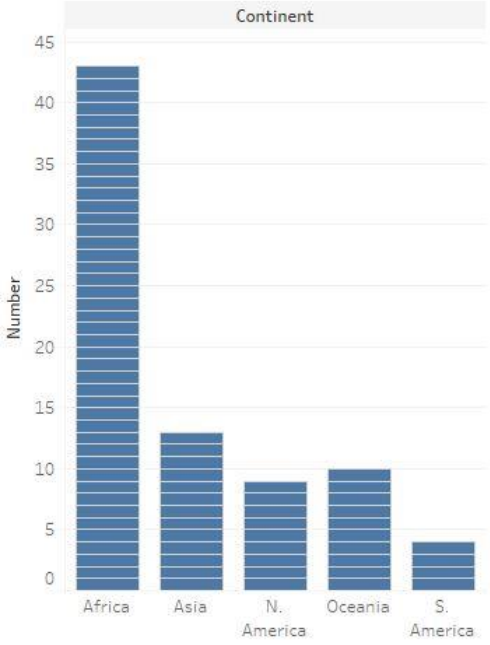
internet users with Avg cost of plans






```sql
-- Is there any countries where less than 50% of population use internet
SELECT continent, count (continent) AS countries_less_than_half_of_population_use_internet
FROM internet_users_cleaned
WHERE percentage_of_population_use_internet <50
Group by continent
```

No.of countries in continent where less than 50% of population use internet



FINDINGS

Around 65% of world population use internet. The analysis shows China, India, US, Indonesia and Brazil are the highest in order of internet users' count. Even though Europe, North and South America have above 70% internet users over population, Asia has more numbers than these continents combined. Asia has highest share of internet users which is more than half of total. There is no significant correlation between number of internet plans and cost of plans.

Europe, Oceania and Asia have more than 39 (Mbit/s)Ookla while rest of the continents between 18-27 (Mbit/s)Ookla The result shows that except Africa, countries vary in internet speed among the continents. Importantly, Asian countries internet speed values have scattered throughout the graph. There is an outstanding difference in cost of internet over the period in countries like China, US, Brazil and Indonesia. However, in 2022, except US, all countries cost of internet comes around same value. The trendline shows the notable relationship between cost of internet and number of internet users. Most surprisingly, 77 countries' internet users are only less than 50% of population.

Here I show the result of analysis (screenshot of Tableau dashboard):