Rajeshwar Vijay Shinde
07th July 2023
Data Source: https://www.kaggle.com/datasets/drahulsingh/largest-malls

PROJECT DOCUMENTATION: ANALYSIS OF WORLD'S LARGEST MALLS DATASET

1. Introduction :-

The purpose of this project is to perform a comprehensive data analysis on the "World's Largest Malls" dataset sourced from **Kaggle**. The dataset provides information about various malls around the world, including their country, city, establishment year, area in square meters, area in square feet, and number of shops. The analysis was conducted using MySQL Workbench and Power BI to uncover valuable insights and trends within the dataset.

2. Methodology :-

2.1 Data Gathering

The "Largest Malls" dataset was obtained from Kaggle (provide link/reference). The dataset consists of a CSV file containing information on mall names, countries, cities, establishment years, area in square meters, area in square feet, and the number of shops. The dataset was imported into MySQL Workbench for further analysis.

2.2 Data Cleaning and Preparation

The following steps were performed to clean and prepare the dataset for analysis:

- Duplicate Removal: Duplicate entries were identified and removed from the dataset to ensure data integrity.

- Missing Values: Missing values were identified and handled appropriately.

- Data Validation: The dataset was validated to ensure the correctness of the data types, constraints, and relationships between columns.

- Data Transformation: Gross area column was split into two separate columns as it contained multiple numeric values of different unit measures.

- Data Import: The CSV file was imported into MySQL Workbench using appropriate table schema and data types.

3. Basic MySQL Data Analysis Questions :-

3.1 What are the distinct countries represented in the dataset?

The analysis revealed the following distinct countries represented in the dataset:

- Iran
- Malaysia
- China
- Philippines
- Thailand
- United States
- Taiwan
- South Korea
- Bangladesh
- Panama
- Indonesia
- Canada
- United Arab Emirates
- Kuwait
- Saudi Arabia
- Russia
- Colombia
- Bahrain
- Venezuela

3.2 What is the total number of malls in each country?

The analysis provided the following count of malls in each country:

country	No_of_malls
Bahrain	1
Bangladesh	1
Canada	1
China	7
Colombia	2
Indonesia	5
Iran	3
Kuwait	1
Malaysia	4
Pakistan	2
Panama	1
Philippines	10
Russia	1
Saudi Arabia	1
South Korea	1
Taiwan	1
Thailand	6
United Ara...	1
United Sta...	5
Venezuela	1

3.3 What is the average area of the malls across each country?

The average area of the malls by country is shown below.

country	AVG_area_mt_sq	AVG_area_sq_ft
Bahrain	273000	2940000
Bangladesh	380000	4100000
Canada	350000	3800000
China	460290	4954287
Colombia	265000	2850000
Indonesia	297893	3208992
Iran	1016667	10933333
Kuwait	425000	4570000
Malaysia	515082	5532500
Pakistan	295000	3200000
Panama	380000	4100000
Philippines	383814	4127936
Russia	258673	2784330
Saudi Arabia	261000	2810000
South Korea	383470	4127600
Taiwan	400000	4300000
Thailand	462546	4978857
United Ara...	350000	3800000
United Sta...	311951	3358600
Venezuela	250000	2700000

3.4 Which mall has the largest area in square feet?

The mall with the largest area in square feet is Iran Mall located in Iran with an area of

21000000 Square feet.

3.5 Which is the oldest large mall in the world?

The oldest large mall in the world is King of Prussia established in 1963.

3.6 How many malls were established in each year?

The number of malls established in each year is as follows:

Year_of_establishment	No_of_malls
1963	1
1967	1
1981	1
1983	1
1985	1
1986	1
1990	1
1991	3
1992	2
1993	1
1995	2
1996	1
1997	2
1998	2
2002	1
2003	1
2004	3
2005	3
2006	3
2007	2
2008	1
2010	3
2011	1
2012	1
2013	4
2014	2
2015	2
2016	2
2017	2
2018	2
2019	1
2022	1

3.7 Which mall has the highest number of shops?

The mall with the highest number of shops is SM Mall of Asia with 3500 shops.

3.8 How many malls have an area greater than 50, 00,000 square meters?

The number of malls with an area greater than 50, 00,000 square meters is 16.

4. Intermediate MySQL Data Analysis Questions :-

4.1 Which city has the most malls?

The city with the most malls is Bangkok with 4 malls.

4.2 How many malls were established in each city?

The number of malls established in each city is as follows:

city	No_of_malls
Bahrain Bay	1
Bangkok	4
Beijing	1
Bloomington, MN (Minneapolis–Saint Paul)	1
Bogota	2
Cagayan de Oro	1
Caracas	1
Cebu City	2
Chengdu	1
Costa Mesa (Greater Los Angeles)	1
Dhaka	1
Dongguan	1
Dubai	1
East Rutherford, NJ (New York City area)	1
Edmonton, Alberta	1
Ekaterinburg	1
Guangzhou	1
Haikou, Hainan	1
Isfahan	1
Jakarta	3
Jeddah	1
Kaohsiung	1
Karachi	1
King of Prussia (Philadelphia metropolitan...)	1
Kuala Lumpur	1
Lahore	1
Makati (Metro Manila)	2
Mandaluyong (Metro Manila)	1
Medan	1
Muntinlupa (Metro Manila)	1
Nonthaburi (Bangkok Metropolitan Region)	1
Panama City	1
Pasay (Metro Manila)	1
Petaling Jaya	1
Phuket	1
Putrajaya	1
Quezon City (Metro Manila)	2
Seoul	1
Shanghai	1
Shiraz	1
Subang Jaya	1
Surabaya	1
Tehran	1
Tianjin	1

4.3 What is the percentage of malls in each country compared to the total?

The percentage of malls in each country compared to the total is as follows:

country	No_of_malls	Percentage_by_country
Bahrain	1	1.82 %
Bangladesh	1	1.82 %
Canada	1	1.82 %
China	7	12.73 %
Colombia	2	3.64 %
Indonesia	5	9.09 %
Iran	3	5.45 %
Kuwait	1	1.82 %
Malaysia	4	7.27 %
Pakistan	2	3.64 %
Panama	1	1.82 %
Philippines	10	18.18 %
Russia	1	1.82 %
Saudi Arabia	1	1.82 %
South Korea	1	1.82 %
Taiwan	1	1.82 %
Thailand	6	10.91 %
United Ara...	1	1.82 %
United Sta...	5	9.09 %
Venezuela	1	1.82 %

4.4 How many malls were established in each decade?

The number of malls established in each decade is as follows:

Decade	No_of_malls
1960	2
1980	4
1990	14
2000	14
2010	20
2020	1

4.5 What is the average number of shops in malls by country?

The average number of shops in malls by country is as follows:

country	AVG_shops
Bahrain	680.0000
Bangladesh	200.0000
Canada	800.0000
China	1032.8571
Colombia	200.0000
Indonesia	332.0000
Iran	1068.3333
Kuwait	1100.0000
Malaysia	588.2500
Pakistan	200.0000
Panama	200.0000
Philippines	833.0000
Russia	250.0000
Saudi Arabia	187.0000
South Korea	200.0000
Taiwan	250.0000
Thailand	375.0000
United Ara...	400.0000
United Sta...	301.2000
Venezuela	300.0000

4.6 Which mall has the highest shop-to-area (meter square) ratio?

The mall with the highest shop-to-area ratio in square meters is SM Mall of Asia.

4.7 How many malls were established in the last five years?

The number of malls established in the last five years is 4.

5. Advanced MySQL Data Analysis Questions :-

5.1 How many malls have an area larger than the average area (m²) of all malls?

The number of malls with an area larger than the average area of all malls is 18.

5.2 What is the average area per shop in each country?

The average area per shop in each country is as follows:

country	area_mt_square	shops	Avg_area_per_shop
Bahrain	273000	680	401.47 m²
Bangladesh	380000	200	1900.00 m²
Canada	350000	800	437.50 m²
China	659612	2350	826.44 m²
Colombia	280000	250	1393.33 m²
Indonesia	360000	150	1194.62 m²
Iran	1950000	2500	1301.58 m²
Kuwait	425000	1100	386.36 m²
Malaysia	821000	650	1073.88 m²
Pakistan	340000	200	1475.00 m²
Panama	380000	200	1900.00 m²
Philippines	589891	3500	765.48 m²
Russia	258673	250	1034.69 m²
Saudi Arabia	261000	187	1395.72 m²
South Korea	383470	200	1917.35 m²
Taiwan	400000	250	1600.00 m²
Thailand	550278	500	1484.18 m²
United Ara...	350000	400	875.00 m²
United Sta...	520257	520	1081.09 m²
Venezuela	250000	300	833.33 m²

5.3 How many malls have an area larger than the average area of malls in their respective countries?

The number of malls with an area larger than the average area of malls in their respective countries is 20 as follows.

country	No_of_malls
China	4
Colombia	1
Indonesia	2
Iran	1
Malaysia	2
Pakistan	1
Philippines	5
Thailand	3
United States	1

5.4 What is the rank of each mall based on its area compared to other malls globally?

The rank of each mall based on its area compared to other malls globally is as follows:

mall_name	area_mt_square	rank_no		mall_name	area_mt_square	rank_no
Iran Mall	1950000	1		Fashion Island (Thaila...	350000	23
IOI City Mall	821000	2		West Edmonton Mall	350000	23
South China Mall	659612	3		The Dubai Mall	350000	23
Isfahan City Center	650000	4		Lucky One Mall	340000	24
SM Mall of Asia	589891	5		Gandaria City	336279	25
SM Tianjin	565000	6		Limketkai Center	320000	26
Golden Resources Mall	557419	7		Berjaya Times Square	320000	26
Central WestGate	550278	8		SM City Fairview	312749	27
CentralWorld	550000	9		Zhengjia Plaza (Gran...	280000	28
ICONSIAM	525000	10		Centro Mayor	280000	28
Mall of America	520257	11		American Dream Mea...	280000	28
1 Utama	519328	12		Haikou International ...	280000	28
SM City North EDSA	497213	13		SM City Cebu	273804	29
Global Harbor	480000	14		The Avenues, Bahrain	273000	30
SM Megamall	474000	15		Medan Centre Point	270000	31
SM Seaside City Cebu	470486	16		Mal Artha Gading	270000	31
Persian Gulf Complex	450000	17		Mall of Arabia	261000	32
The Avenues Mall	425000	18		King of Prussia	259500	33
Sunway Pyramid	400000	19		Greenwich Mall	258673	34
New Century Global ...	400000	19		Tunjungan Plaza	253187	35
Dream Mall	400000	19		Emporium Mall	250000	36
Siam Paragon	400000	19		Centro Sambil	250000	36
Central Phuket	400000	19		Aventura Mall	250000	36
Festival Alabang	400000	19		Glorietta	250000	36
Lotte World Mall	383470	20		Greenbelt	250000	36
Jamuna Future Park	380000	21		South Coast Plaza	250000	36
Albrook Mall	380000	21		Centro Comercial San...	250000	36
Mal Taman Anggrek	360000	22				

5.5 Which country has the 2nd highest number of total shops inside the malls?

The country with the second-highest number of total shops inside the malls is China.

5.6 How many pairs of malls have similar areas but are located in different countries?

There are total 43 pairs of malls have similar areas but are located in different countries

6. Power BI Report :-

To visually explore and communicate the insights derived from the "World's Largest Malls " dataset, a Power BI report was created. The report includes various visualizations and interactive features to enhance data exploration and understanding. Below are the key components and insights presented in the Power BI report:

1. Data Source: The "World's Largest Malls" dataset was imported into Power BI as the primary data source for the report.

2. Data Modeling: The dataset was transformed and modeled in Power BI using appropriate relationships and data types to enable seamless data visualization.

3. Dashboard: A visually appealing and informative dashboard was created as the main landing page of the Power BI report. It provides an overview of the key metrics and insights derived from the dataset.

4. Visualizations: A wide range of visualizations were employed to present the insights derived from the data analysis. These include:

- o Cards: Utilized to display key metrics such as the largest mall area, the oldest mall, and the highest shop-to-area ratio.
- o Map: Used to showcase the geographical distribution of malls by country and city.
- o Pie Chart: Employed to illustrate the shops to area ration with respect to each country.
- o Decomposition Tree: Utilized to analyze and visualize hierarchical relationships within the data, enabling exploration of factors influencing mall characteristics.
- o Line and Clustered Column Chart: Used to demonstrate the trend of no. of malls and shop to area ratio by country.
- o Tree map: Employed to provide an overview of shop distribution based on country.
- o Table: Used to present detailed information, including mall names, countries, cities, establishment years.
- o Clustered Bar Chart: Employed to compare the total number of malls in each country.
- o Clustered Column Chart: Utilized to compare the number of malls in different decades, providing insights into the distribution of mall establishments over time.
- o Area Chart: Used to visualize the trend of mall area by country.

These visualizations were carefully selected to effectively communicate the insights derived from the data analysis. They enhance the understanding of patterns, trends, and comparisons within the dataset, enabling users to gain valuable insights and make informed decisions based on the findings.

5. Interactivity: The Power BI report provides interactive features such as filtering, highlighting, and drill-through options to allow users to explore the data further and gain more specific insights based on their interests.

6. Insights and Findings: The Power BI report showcases the key findings from the data analysis process, answering the SQL questions and providing a visual representation of the derived insights. These insights include the distribution of malls by country, the average mall area, the cities with the most malls, the relationship between the number of shops and the mall area, and more.

The Power BI report adds an interactive and visually appealing dimension to your project, allowing users to interact with the data and explore the insights in a user-friendly manner.

7. Conclusion

The project documentation demonstrates a comprehensive analysis of the "World's Largest Malls" dataset. By combining MySQL Workbench for data analysis and Power BI for data visualization, valuable insights and trends were extracted from the dataset. The documentation and the accompanying Power BI report provide a clear understanding of the dataset, the analysis methodology, and the derived insights, contributing to a comprehensive and well-documented data analysis project.