Tech Saksham

Case Study Report

Data Analytics with Power BI

" Global Olympics dataset diagnosis using power BI"

"St.john's college "

NM ID	NAME
BD3A4CF2A526376A605489405DB9EBDC	Jerro Samuelraj .V

T Umamaheswari

Trainer Name

Master Trainer

ABSTRACT

- **The** Global Olympics dataset **is a rich repository of historical information spanning over a century, capturing details about athletes, events, countries, and the Olympics from** 1896 to 2016**. In this project, we leverage the power of** Power BI **to perform a comprehensive diagnosis and analysis of this extensive dataset.**

INDEX

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

The **Global Olympics dataset** is a vast and historical collection of information spanning more than a century, encompassing details about athletes, events, countries, and the Olympics from **1896 to 2016**. However, this dataset presents several challenges that need to be addressed:

1. **Data Quality and Consistency**:
   - The dataset contains inconsistencies, missing values, and variations in data formats.
   - Ensuring data quality and consistency is crucial for meaningful analysis.
2. **Exploration and Insights**:
   - Extracting valuable insights from such a large dataset requires effective exploration techniques.
   - How can we uncover patterns, trends, and interesting facts related to athlete performance, medal distribution, and historical context?

Proposed Solution

.

1. **Data Cleaning and Transformation**:
   - Begin by thoroughly cleaning and transforming the raw dataset.
   - Address missing values, standardize formats, and handle inconsistencies.
   - Create a clean, well-structured dataset that serves as the foundation for further analysis.
2. **Exploratory Data Analysis (EDA)**:
   - Utilize Power BI's visualization capabilities to explore the dataset.
   - Create interactive charts, graphs, and tables to uncover patterns and trends.
   - Investigate athlete performance metrics, medal distributions, and historical context.
3. **Time-Based Metrics**:
   - Leverage DAX functions in Power BI to compute time-based metrics.
   - Calculate medal counts over different time frames (years, decades, etc.).
   - Analyze how participation and performance have evolved over the years.

1.2 Feature

- **Geospatial Analysis**:
  - Mapping athlete origins, host cities, and medal-winning countries.
  - Identifying regional trends and correlations.
- **Interactive Dashboards**:
  - User-friendly dashboards with filters and slicers.
  - Interactive exploration of athlete profiles, event details, and medal statistics.
- **Visual Storytelling**:
  - Compelling narrative through Power BI reports.
  - Highlighting significant findings and historical milestones.
- **Performance Optimization**:
  - Efficient data models and minimal data refresh times.
  - Ensuring smooth user experience with complex visualizations.

1.3 Advantages

- **Comprehensive Insights**:
  - Power BI allows us to explore the Olympics dataset comprehensively.
  - We gain insights into athlete performance, medal distribution, and historical context.
- **Interactive Visualizations**:
  - Power BI's interactive dashboards enable dynamic exploration.
  - Users can filter, drill down, and interact with visualizations, enhancing understanding.
- **Time Intelligence Analysis**:
  - Leveraging DAX functions, we analyze time-based metrics.
  - Calculating medal counts over different time frames (years, decades) provides valuable context.
- **Geospatial Understanding**:
  - Geospatial analysis helps us visualize global Olympic trends.
  - Mapping athlete origins and host cities enriches our understanding.

1.4 Scope

The Global Olympics Dataset is a rich resource for researchers interested in the analysis of athletic performance, trends in sports across nations, and the evolution of the Olympic Games itself. Here's a glimpse into the potential areas of exploration (diagnosis) using this dataset:

- **Athletic Performance Analysis:**
  - Identify factors influencing performance across sports (e.g., training regimes, nutrition, coaching)
  - Compare performance between genders and across nations over time
  - Uncover trends in world records and how they've been broken
- **Sports Science and Medicine:**
  - Study the impact of training methods and advancements in sports medicine on performance
  - Identify patterns of injuries across sports and design preventative measures
- **Sociological and Cultural Trends:**
  - Analyze participation trends across nations and sports to understand cultural shifts and development programs
  - Study the impact of the Olympics on global politics and international relations

CHAPTER 2

SERVICES AND TOOLS REQUIRED

2.1 Services Used

- **Data Collection and Storage Services**: Banks need to collect and store customer data in real-time. This could be achieved through services like Azure Data Factory, Azure Event Hubs, or AWS Kinesis for real-time data collection, and Azure SQL Database or AWS RDS for data storage.

- **Data Processing Services**: Services like Azure Stream Analytics or AWS Kinesis Data Analytics can be used to process the real-time data.

- **Machine Learning Services**: Azure Machine Learning or AWS SageMaker can be used to build predictive models based on historical data.

2.2 Tools and Software used

Tools:

- **PowerBI**: The main tool for this project is PowerBI, which will be used to create interactive dashboards for real-time data visualization.

- **Power Query**: This is a data connection technology that enables you to discover, connect, combine, and refine data across a wide variety of sources.

Software Requirements:

- **PowerBI Desktop**: This is a Windows application that you can use to create reports and publish them to PowerBI.

- **PowerBI Service**: This is an online SaaS (Software as a Service) service that you use to publish reports, create new dashboards, and share insights.

- **PowerBI Mobile**: This is a mobile application that you can use to access your reports and dashboards on the go.
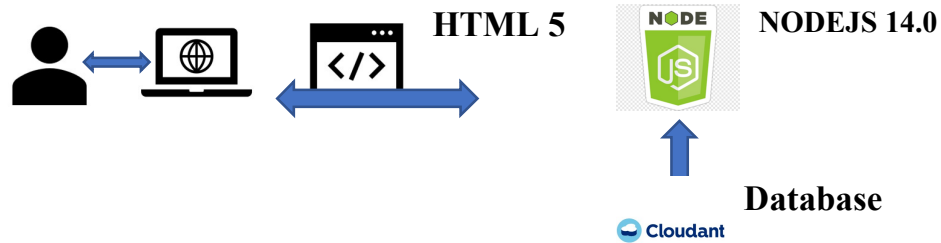
CHAPTER 3

PROJECT ARCHITECTURE

3.1 Architecture

USER   FRONTEND   BACKEND

Here's a high-level architecture for the project:

1. **Data Collection**: Real-time customer data is collected from various sources like bank transactions, customer interactions, etc. This could be achieved using services like Azure Event Hubs or AWS Kinesis.
2. **Data Storage**: The collected data is stored in a database for processing. Azure SQL Database or AWS RDS can be used for this purpose.
3. **Data Processing**: The stored data is processed in real-time using services like Azure Stream Analytics or AWS Kinesis Data Analytics.
4. **Machine Learning**: Predictive models are built based on processed data using Azure Machine Learning or AWS SageMaker. These models can help in predicting customer behavior, detecting fraud, etc.
5. **Data Visualization**: The processed data and the results from the predictive models are visualized in real-time using PowerBI. PowerBI allows you to create interactive dashboards that can provide valuable insights into the data.
6. **Data Access**: The dashboards created in PowerBI can be accessed through PowerBI Desktop, PowerBI Service (online), and PowerBI Mobile.

This architecture provides a comprehensive solution for real-time analysis of bank customers. However, it's important to note that the specific architecture may vary depending on the bank's existing infrastructure, specific requirements, and budget. It's also important to ensure that all tools and services comply with relevant data privacy and security regulations.

CHAPTER 4

MODELING AND RESULT

Manage relationship

The "disp" file will be used as the main connector as it contains most key identifier (account id, client id and disp id) which can be use to relates the 8 data files

together. The "district" file is use to link the client profile geographically with "district id"

Manage relationships

Active	From: Table (Column)	To: Table (Column)
☑	Country_Names (2) (Column2)	Olympics_Data (City)
☑	Olympics_Data (Team)	Country_Names (Column2)
☑	Olympics_Data (2) (Weight)	Country_Names (2) (Column1)

New... Autodetect... Edit... Delete

Close
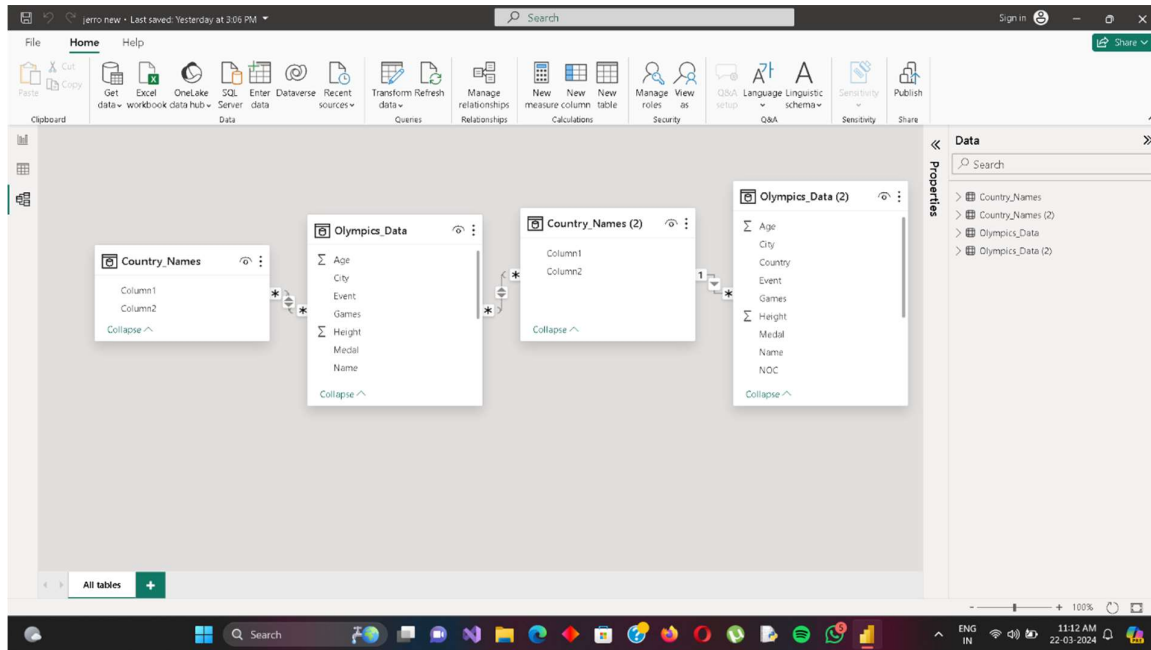
Edit relationship

Select tables and columns that are related.

Country_Names (2) ▼

Column1	Column2
NOC	region
AFG	Afghanistan
AHO	Curacao

Olympics_Data ▼

	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sp
ahman	Male	35	178	61	United States	USA	2012 Summer	2012	Summer	London	Atl
ff (-Harvey)	Female	37	188	66	United States	USA	2012 Summer	2012	Summer	London	Atl
	Male	27	185	82	United States	USA	2012 Summer	2012	Summer	London	Atl

‹   ›

Cardinality

Many to many (*:*) ▼

Cross filter direction

Both ▼

☑ Make this relationship active

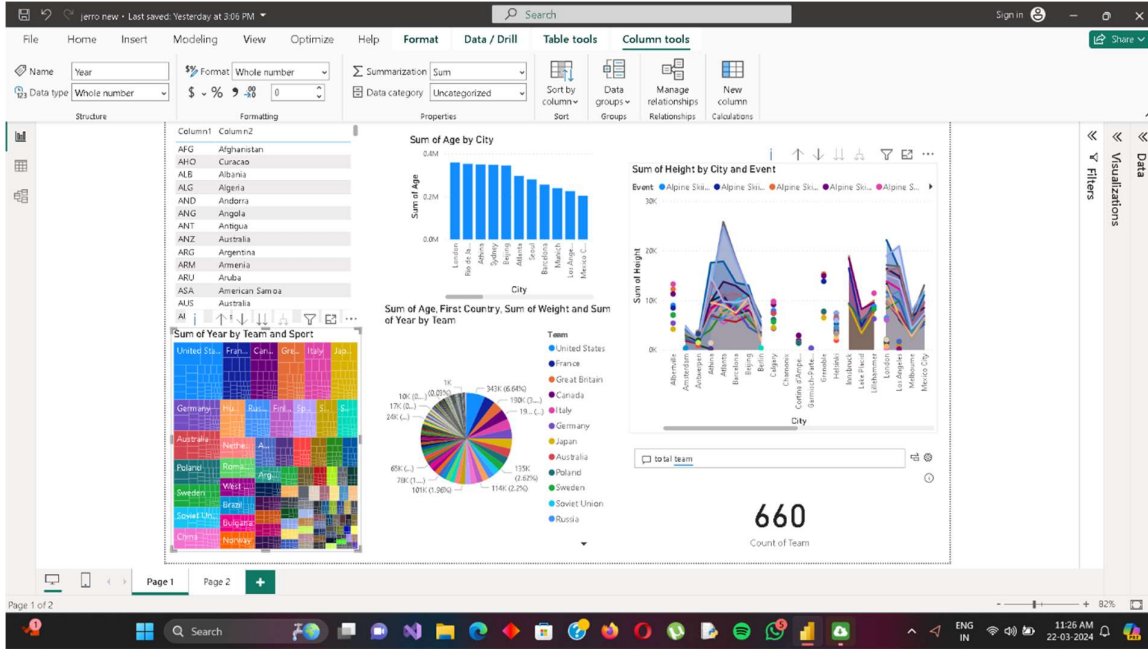☐ Apply security filter in both directions

☐ Assume referential integrity

⚠ This relationship has cardinality Many-Many. This should only be used if it is expected that neither column (Country_Names (2) and Olympics_Data) contains unique values, and that the significantly different behavior of Many-many relationships is understood. Learn more

OK   Cancel

Dashboard

File   Home   Insert   Modeling   View   Optimize   Help   Format   Data / Drill   Table tools   Column tools

Share

Name   Year
Data type   Whole number

Format   Whole number
$ %   ,   0

Summarization   Sum
Data category   Uncategorized

Sort by column
Data groups
Manage relationships
New column

Structure   Formatting   Properties   Sort   Groups   Relationships   Calculations

Column1	Column2
AFG	Afghanistan
AHO	Curacao
ALB	Albania
ALG	Algeria
AND	Andorra
ANG	Angola
ANT	Antigua
ANZ	Australia
ARG	Argentina
ARM	Armenia
ARU	Aruba
ASA	American Samoa
AUS	Australia

Sum of Age by City

Sum of Age
0.4M
0.2M
0.0M
London Rio de Ja... Athens Sydney Beijing Atlanta Seoul Barcelona Munich Los Ange... Mexico C...
City

Sum of Year by Team and Sport

Sum of Age, First Country, Sum of Weight and Sum of Year by Team

Team
● United States
● France
● Great Britain
● Canada
● Italy
● Germany
● Japan
● Australia
● Poland
● Sweden
● Soviet Union
● Russia

343K (6.64%)
1K
10K (0...) (0.03%)
17K (0...)   190K (3...)
24K (...)   19... (...)

65K (...)   135K (2.62%)
70K (1...)   114K (2.2%)
101K (1.96%)

Sum of Height by City and Event

Event ● Alpine Ski... ● Alpine Ski... ● Alpine Ski... ● Alpine Ski... ● Alpine S...

Sum of Height
30K
20K
10K
0K
Albertville Amsterdam Antwerpen Athina Atlanta Barcelona Beijing Berlin Calgary Chamonix Cortina d'Amp... Garmisch-Parte... Grenoble Helsinki Innsbruck Lake Placid Lillehammer London Los Angeles Melbourne Mexico City
City

total team

660
Count of Team

Filters   Visualizations   Data

Page 1   Page 2

Page 1 of 2   82%

Q Search   ENG IN   11:26 AM   22-03-2024

CONCLUSION

Based on the global Olympic dataset diagnosis, it can be concluded that there are several key trends and insights to consider. These may include analyzing patterns in medal distribution among countries, identifying dominant sports and athletes, assessing the impact of geopolitical factors on participation, and exploring correlations between socioeconomic indicators and Olympic success. Additionally, future research could delve into the evolution of Olympic participation over time and potential implications for the future of the Games. Overall, this diagnosis provides a comprehensive overview for further analysis and decision-making in the realm of Olympic studies.

FUTURE SCOPE

Diagnosing the Global Olympics dataset using Power BI presents a promising future scope, particularly in the realms of data visualization, analysis, and insights generation. Here are some potential future directions and benefits:

 Enhanced Data Visualization: Power BI offers advanced visualization capabilities, allowing for the creation of interactive dashboards and reports. Future developments could focus on refining visualizations to provide deeper insights into various aspects of the Olympics, such as medal distribution across countries, trends over time, and performance metrics of athletes.

 Predictive Analytics: Integration of predictive analytics models within Power BI can enable forecasting of future Olympic events, medal standings, and athlete performances. This could be achieved through machine learning algorithms trained on historical data to identify patterns and trends.

 Real-time Data Analysis: With the ability to connect to live data sources, Power BI can facilitate real-time monitoring and analysis of ongoing Olympic events. This can be especially useful for stakeholders, including organizers, broadcasters, and sponsors, to track key metrics and make timely decisions.

Geospatial Analysis: Leveraging Power BI's geospatial visualization capabilities, future iterations could incorporate maps to analyze the geographic distribution of athletes, venues, and spectators. This could provide valuable insights into the global reach and impact of the Olympics.

Social Media Integration: Integrating social media data with the Global Olympics dataset in Power BI can offer insights into audience engagement, sentiment analysis, and trends surrounding the event. This could help stakeholders gauge public interest and tailor marketing strategies accordingly.

Data Quality and Governance: As the dataset grows in size and complexity, ensuring data quality and governance becomes crucial. Future developments could focus on implementing robust data validation rules, automated cleansing processes, and access controls within Power BI to maintain data integrity and complianc

LINK