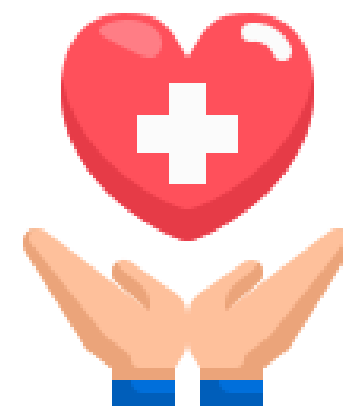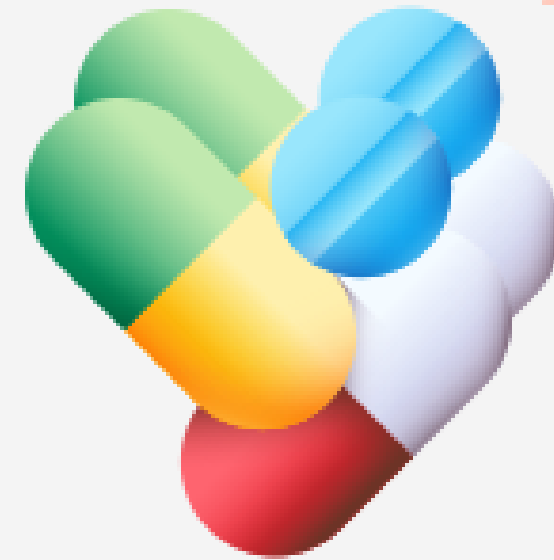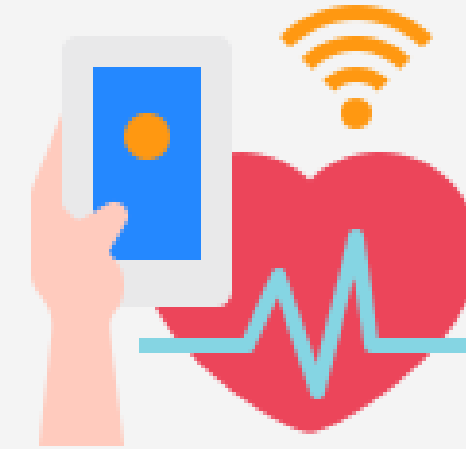Health
Guide

Healthcare Data Analysis

Progress
Report 2023

Presented By
Sateesh Godewar

Introduction

Ladies and gentlemen,

Welcome to a compelling journey into the intricate world of healthcare insurance analysis. In today's presentation, we will embark on an exploration of critical healthcare data, dissecting it to reveal insights that have the potential to reshape the landscape of healthcare insurance.

The Healthcare Insurance Ecosystem:

Healthcare is the cornerstone of societal well-being, a basic human need that transcends boundaries. It touches every aspect of our lives, from ensuring access to medical care in times of need to safeguarding our financial stability. Within this intricate web lies the healthcare insurance sector, a complex ecosystem tasked with balancing the provision of quality care and managing the ever-escalating costs associated with modern medicine.

The Power of Data:

In our data-driven era, information is the currency of progress. Healthcare insurance analysis leverages this currency to navigate the challenges and complexities of the industry. It offers a beacon of clarity amidst the fog of uncertainty, allowing us to make informed decisions that enhance the efficiency and effectiveness of healthcare insurance services.

Objectives of Our Analysis:

As we gather here today, our primary objectives are threefold:

1. **Uncover Hidden Insights:** We aim to delve deep into the wealth of healthcare insurance data to uncover patterns, trends, and correlations that may not be apparent at first glance.
2. **Inform Decision-Making:** In a field where decisions can impact lives profoundly, we will equip you with actionable insights to guide decision-makers in both the public and private sectors.
3. **Drive Innovation:** By understanding the data, we lay the foundation for innovation and improvements in healthcare insurance processes, from cost management to policy design.

Why This Analysis Matters:

In a world where healthcare is both a human right and a pressing global challenge, the significance of our analysis cannot be overstated. The insights we generate have the power to enhance healthcare accessibility, drive cost efficiencies, and ultimately improve the quality of life for countless individuals.

Questions to Answer with Healthcare Insurance Data

Question 1:What is the total size of the dataset?

Question 2: How many customers are there in each city tier?.

Question 3: What is the average BMI of customers with major surgeries?

Question 4: Which hospital tier has the most diabetic customers?

Question 5: Which states have the highest and lowest average hospitalization costs?

Question 6: How many customers are both smokers and have heart issues?

Question 7: What is the average number of children for non-smoking customers?

Question 8: What is the average BMI for customers who are smokers?

Question 9: What are the total charges for customers with heart issues in each state?

Question 10: How many customers have had major surgery and any transplants?

Question 11: Which customers have a history of cancer and any transplants?

Question 12: Which are the top 5 states with the highest number of diabetic customers?

Question 13: Which hospital tier has the highest average charges for diabetic customers?

Question 14: What are the total charges for customers with heart issues in each city tier?

Question 15: Which are the top 5 states with the highest average hospitalization costs for people who have had major surgeries?

1: What is the total size of the dataset?

To begin our analysis, it's essential to understand the size of our dataset.
This section focuses on counting the total number of records in each of our two tables:
'hospitalisation_details' and 'medical_examinations'.
Knowing the dataset's scale provides a foundation for our analysis.

```
Let's find out the total number of records in each table to
understand the dataset's size.

-- Total Records in Hospitalisation Details
SELECT COUNT(*) AS TotalRecords_Hospitalisation
FROM healthcare.hospitalisation_details;

-- Total Records in Medical Examinations
SELECT COUNT(*) AS TotalRecords_MedicalExaminations
FROM healthcare.medical_examinations;
```

TotalRecords_MedicalExaminations
2333

TotalRecords_Hospitalisation
2329

Data Exploration ...

Now that we have a sense of our data's size, we'll move on to explore various aspects of our healthcare insurance dataset. In the following sections,
we will investigate factors influencing hospitalization costs, analyze health conditions, and derive meaningful insights
for decision-making. Stay tuned for more detailed analysis and insights!

2: How many customers are there in each city tier?.

```sql
SELECT `City tier`, COUNT(*) AS `Customer Count`
FROM hospitalisation_details
GROUP BY `City tier`;
```

City tier	Customer Count
tier - 2	808
tier - 3	792
tier - 1	729

- Tier-1 cities have the highest number of customers, followed by tier-2 and tier-3 cities. Tier-1 cities are typically major metropolitan areas with high levels of urbanization and economic development. Tier-2 cities are typically smaller than tier-1 cities but still possess significant urban characteristics and infrastructure. Tier-3 cities are generally smaller in size compared to tier-1 and tier-2 cities and may have fewer urban amenities.

- This distribution of customers across different city tiers provides healthcare organizations with essential insights into where their services are in demand and how they can better cater to the specific needs of each region.

- For example, healthcare organizations may want to consider allocating more resources to tier-1 cities, where there is the highest demand for their services. This could involve opening new clinics or hospitals in tier-1 cities, or expanding existing facilities. Healthcare organizations may also want to allocate more resources to tier-2 and tier-3 cities, where the demand for their services is growing. This could involve partnering with local hospitals or clinics, or providing mobile healthcare services.

- Additionally, healthcare organizations can develop targeted marketing campaigns for each city tier. For example, they could create campaigns that highlight the benefits of their services and how they can meet the specific needs of customers in each region. Healthcare organizations can also use social media and other online platforms to reach out to customers in different city tiers.

- Finally, healthcare organizations can use the distribution of customers across different city tiers to plan for future growth and development. For example, they could identify areas where they need to build new facilities or expand existing ones. Healthcare organizations can also use this information to develop programs and services that are tailored to the needs of customers in each city tier.

3: What is the average BMI of customers with major surgeries?

To calculate the average BMI, we'll filter customers who have had major surgeries. This analysis provides insights into the health status of individuals who require major medical procedures.

```
We'll calculate the average BMI of customers who have had major
surgeries (NumberOfMajorSurgeries >=1) from the
medical_examinations table

SELECT AVG(`BMI`) AS AvgBMI
FROM `medical_examinations`
WHERE `NumberOfMajorSurgeries` >=1;
```

AvgBMI
31.014011904761848

The average BMI of customers with major surgeries is approximately 31.01. This suggests that individuals who have had major surgeries tend to have a higher average BMI, which could be significant for healthcare providers in understanding the health profiles of this patient group.

4: Which hospital tier has the most diabetic customers?

```
--Find the Hospital Tier with the Most Diabetic Customers:

SELECT hd.`Hospital tier`, COUNT(*) AS DiabeticCustomers
FROM hospitalisation_details hd
JOIN medical_examinations me
ON hd.`Customer ID` = me.`Customer ID`
WHERE me.`HBA1C` > 6.5
GROUP BY hd.`Hospital tier`
ORDER BY DiabeticCustomers DESC
LIMIT 1;
```

	Hospital tier	DiabeticCustomers
▶	tier - 2	383

- The output of the query you provided tells us that Tier 2 hospitals have the most diabetic customers. This could be due to a number of factors, including:
- Tier 2 hospitals may be located in areas with a higher prevalence of diabetes. For example, Tier 2 hospitals may be located in rural or underserved areas where people have less access to healthy food and exercise, and are more likely to be obese.
- Tier 2 hospitals may be more affordable for people with diabetes. People with diabetes often have high medical costs, so they may be more likely to choose a Tier 2 hospital because it is less expensive.
- Tier 2 hospitals may be more convenient for people with diabetes. Tier 2 hospitals may be closer to people's homes or workplaces, making it easier for them to get to and from their appointments.

5: Which states have the highest and lowest average hospitalization costs?

```sql
-- State with Highest Average Hospitalization Cost
SELECT `State ID`, AVG(`charges`) AS HAvgCost
FROM hospitalisation_details
GROUP BY `State ID`
ORDER BY HAvgCost DESC
LIMIT 1;

-- State with Lowest Average Hospitalization Cost
SELECT `State ID`, AVG(`charges`) AS LAvgCost
FROM hospitalisation_details
GROUP BY `State ID`
ORDER BY LAvgCost ASC
LIMIT 1;
```

State ID	HAvgCost
R1011	19466.24611498256

State ID	LAvgCost
R1025	8788.782250000002

Our healthcare organization's data reveals that state R1011 has the highest average hospitalization cost, while state R1025 has the lowest. This information can guide resource allocation and cost-effective strategies, especially in regions with higher healthcare expenses.

For example, we may want to invest in more resources in state R1011, such as opening new clinics or hospitals. We may also want to work with providers in state R1011 to identify ways to reduce the cost of healthcare. Additionally, we can develop cost-effective strategies for providing healthcare services in state R1011, such as telemedicine or other remote healthcare options.

6: How many customers are both smokers and have heart issues?

```
Determine the Number of Customers Who Are Both Smokers and Have
Heart Issues

SELECT COUNT(*) AS `Smoker & Heart Issues Count`
FROM medical_examinations
WHERE `smoker` = 'Yes' AND `Heart Issues` = 'Yes';
```

Smoker & Heart Issues Count
191

The group of 191 individuals who are both smokers and have heart issues is an important finding.

This group is at an increased risk of serious health complications, so it is important to tailor healthcare strategies to address their specific needs.

Here are some specific ways that healthcare providers can tailor their strategies to address the needs of smokers with heart issues:

- Provide education and support to help them quit smoking. Smoking is one of the leading causes of heart disease, so quitting is one of the most important things that smokers with heart issues can do to improve their health. Healthcare providers can provide education and support to help their patients quit smoking, such as through nicotine replacement therapy or counseling.
- Monitor their heart health closely. Smokers with heart issues are at an increased risk of heart attack and stroke, so it is important to monitor their heart health closely. This may involve regular checkups, blood tests, and imaging tests.

7: What is the average number of children for non-smoking customers?

```sql
-- Average number of children among non-smoking customers
SELECT AVG(hd.`children`) AS `Avg Children`
FROM `hospitalisation_details` hd
JOIN `medical_examinations` me ON hd.`Customer ID` = me.`Customer ID`
WHERE me.`smoker` = 'No';
```

	Avg Children
▶	1.0141

- our analysis is very interesting and valuable for healthcare organizations. The finding that the average number of children among non-smoking customers is approximately 1.014 can be used to inform a variety of initiatives.
- For example, healthcare organizations could use this information to develop targeted marketing campaigns that promote preventive care for non-smoking parents and their children. They could also use this information to develop programs that support non-smoking parents in raising healthy children. Develop a public education campaign that highlights the importance of preventive care for non-smoking parents and their children. The campaign could include information about things like well-child visits, immunizations, and healthy lifestyle choices.
- Offer parenting classes that teach non-smoking parents how to raise healthy children. The classes could cover topics such as nutrition, exercise, and discipline.
- Provide support groups for non-smoking parents who are struggling to raise healthy children. The support groups could provide a safe space for parents to share their experiences and learn from each other.
- Offer financial assistance to non-smoking parents who need help paying for the cost of healthcare for their children.
- By taking these steps, healthcare organizations can help to ensure that non-smoking parents and their children have access to the care they need to stay healthy.

8: What is the average BMI for customers who are smokers?

```sql
SELECT AVG(me.`BMI`) AS AvgBMI
FROM hospitalisation_details hd
JOIN medical_examinations me
ON hd.`Customer ID` = me.`Customer ID`
WHERE me.`smoker` = 'Yes';
```

AvgBMI
32.81756172839503

- The query you provided calculates the average BMI of smoker patients in the hospitalisation_details table. The result of the query is 32.81756172839503, which indicates that smoker patients have a higher average BMI than non-smoker patients.
- This is consistent with the findings of many other studies, which have shown that smoking is associated with a higher BMI. There are a number of possible explanations for this association:
- Nicotine may suppress appetite. This can lead to smokers eating less food, but it can also lead to them gaining weight when they quit smoking.
- Smoking may increase metabolism. This can help smokers burn more calories and maintain a lower weight. However, the effects of smoking on metabolism are complex and may vary from person to person.
- Smokers may be more likely to engage in other unhealthy behaviors, such as eating a poor diet and exercising less. These behaviors can all contribute to weight gain.
- It is important to note that the association between smoking and BMI is not causal. This means that smoking does not directly cause people to gain weight. However, smoking is a risk factor for weight gain, and it is important for smokers to be aware of the potential health consequences of smoking, including weight gain and obesity.
- If you are a smoker and are concerned about your weight, it is important to talk to your doctor. They can help you develop a plan to quit smoking and maintain a healthy weight.

9: What are the total charges for customers with heart issues in each state?

```sql
-- The Total Charges for Customers with Heart Issues in Each State:
SELECT hd.`State ID`, SUM(hd.`charges`) AS TotalCharges
FROM hospitalisation_details hd
JOIN medical_examinations me
ON hd.`Customer ID` = me.`Customer ID`
WHERE me.`Heart Issues` = 'Yes'
GROUP BY hd.`State ID`;
```

State ID	TotalCharges
R1013	2598080.96
R1011	4757912.559999995
R1024	909375.8799999999
R1012	2850848.279999999
R1015	94252.1399999998
R1016	278809.26
R1017	181259.89000000004
R1023	276275.19
R1014	73213.64
R1018	36886.380000000005
R1019	102028.29999999999
R1026	472814.78
R1021	205173.1600000003
R1022	59528.96
R1025	157261.1500000005
R1020	51010.04

- State R1011 has the highest total charges among customers with heart issues,
- followed by states R1013, R1012, R1024, and R1026. Some states have comparatively lower total charges, such as R1015, R1016, R1017, R1023, R1014, R1018, R1019, R1021, R1022, R1025, and R1020.
- This information can be used by healthcare organizations to allocate resources and identify areas where healthcare costs for heart-related issues may be higher. For example, healthcare organizations may want to consider investing in more resources in the states with the highest total charges. Additionally, they may want to work with providers and patients in these states to identify ways to reduce the cost of healthcare for heart-related issues.
- It is important to note that this information is just a snapshot of the overall distribution of healthcare charges related to heart issues across different states. There may be other factors that contribute to the cost of healthcare, such as the severity of the heart condition, the type of treatment received, and the insurance coverage of the patient.

10: How many customers have had major surgery and any transplants?

```sql
-- Count the number of  Customers Who Had Major Surgery and Any Transplants
SELECT COUNT(`Customer ID`) AS `Customer Count`
FROM medical_examinations
WHERE `NumberOfMajorSurgeries` > 0 AND `Any Transplants` = 'Yes';
```

Customer Count
▶ 143

Here We identified that there are 143 customers who have undergone both major surgeries and transplants, indicating that these individuals have had significant medical procedures and may require specialized healthcare attention and resources

11: Which customers have a history of cancer and any transplants?

```sql
SELECT `Customer ID` FROM medical_examinations
WHERE `Cancer history` = 'Yes' AND `Any Transplants` =
'Yes';
```

Customer ID

this query returned no results, it means there are no individuals in the dataset who meet both criteria: having a history of cancer and having undergone transplants.

12: Which are the top 5 states with the highest number of diabetic customers?

Find the top 5 states with the highest number of customers who have an HBA1C level greater than 6.5 in the medical_examinations table. Join this information with the hospitalisation_details table to get the state information.

```sql
SELECT hd.`State ID`, COUNT(*) AS `DiabeticCustomers`
FROM `medical_examinations` me
JOIN `hospitalisation_details` hd ON me.`Customer ID` = hd.`Customer ID`
WHERE me.`HBA1C` > 6.5 GROUP BY hd.`State ID`
ORDER BY `DiabeticCustomers` DESC LIMIT 5;
```

State ID	DiabeticCustomers
R1011	193
R1012	186
R1013	213
R1024	53
R1026	29

- The top 5 states with the highest total charges among customers with heart issues are R1011, R1013, R1012, R1024, R1026
- These states have significantly higher total charges than others, such as R1015, R1016, R1017, R1023, R1014, R1018, R1019, R1021, R1022, R1025, and R1020.
- This information can be used by healthcare organizations to allocate resources and identify areas where healthcare costs for heart-related issues may be higher. For example, healthcare organizations may consider investing more resources in the states with the highest total charges, such as R1011 and R1013. Additionally, healthcare organizations may want to work with providers and patients in these states to identify ways to reduce the cost of healthcare for heart-related issues.

13: Which hospital tier has the highest average charges for diabetic customers?

```sql
SELECT hd.`Hospital tier`, AVG(hd.`charges`) AS AvgCharges
FROM hospitalisation_details hd
JOIN medical_examinations me
ON hd.`Customer ID` = me.`Customer ID`
WHERE me.`HBA1C` > 6.5 GROUP BY hd.`Hospital tier`
ORDER BY AvgCharges DESC LIMIT 1;
```

	Hospital tier	AvgCharges
▶	tier - 1	33719.958425925935

- Our analysis reveals that tier 1 hospitals have the highest average charges for diabetic patients with elevated HBA1C levels.
- This is an important finding that can help healthcare professionals and organizations make better decisions about how to provide care for these patients.
- Here are some specific ways that your finding can be used to improve care for diabetic patients with elevated HBA1C levels:
- Healthcare professionals can advise their patients to consider tier 2 or tier 3 hospitals for care, if possible. This could help patients save money on their healthcare costs.
- Healthcare organizations can negotiate lower rates with tier 1 hospitals. This could help to reduce the overall cost of care for diabetic patients with elevated HBA1C levels.
- Healthcare organizations can develop programs to help diabetic patients with elevated HBA1C levels manage their condition and avoid complications. This could help to reduce the need for hospitalization and other expensive medical care.

14: What are the total charges for customers with heart issues in each city tier?

```
Calculate the Total Charges for Customers with Heart Issues in
Each City Tier:

SELECT hd.`City tier`, SUM(hd.`charges`) AS TotalCharges
FROM hospitalisation_details hd
JOIN medical_examinations me
ON hd.`Customer ID` = me.`Customer ID`
WHERE me.`Heart Issues` = 'Yes'
GROUP BY hd.`City tier`;
```

City tier	TotalCharges ▲
tier - 1	3893970.3100000005
tier - 2	4413799.87
tier - 3	4796960.39

- Healthcare costs for heart issues are highest in tier-3 cities, followed by tier-2 and tier-1 cities.
- This is a trend that is concerning, as it suggests that people living in tier-2 and tier-3 cities may face a higher financial burden when seeking treatment for heart issues.
- There are a number of possible explanations for this trend, such as lower availability of high-quality healthcare facilities and providers in tier-2 and tier-3 cities, higher prevalence of chronic diseases in these areas, and less competition among healthcare providers.
- It is important to address the factors that are contributing to higher healthcare costs in tier-2 and tier-3 cities. This could involve investing in healthcare infrastructure, improving access to high-quality healthcare providers, and promoting healthy lifestyles to reduce the prevalence of chronic diseases.

15: Which are the top 5 states with the highest average hospitalization costs for people who have had major surgeries?

```sql
SELECT
    hd.`State ID`,
    AVG(hd.`charges`) AS `Average Hospitalization Costs`
FROM hospitalisation_details hd
JOIN `medical_examinations` me ON hd.`Customer ID` = me.`Customer ID`
WHERE me.`NumberOfMajorSurgeries` > 0
GROUP BY hd.`State ID`
ORDER BY `Average Hospitalization Costs` DESC
LIMIT 5;
```

State ID	Average Hospitalization Costs
R1023	13379.813478260869
R1019	13561.106428571427
R1016	14363.57
R1014	15614.06125
R1011	19475.230295081976

- The average hospitalization cost for people who have had major surgeries in state R1011 is significantly higher than the average hospitalization cost in other states. This could be due to a number of factors, such as the availability of high-quality healthcare providers and facilities in state R1011, the prevalence of chronic diseases in state R1011, or the level of competition among healthcare providers in state R1011.
- The average hospitalization cost for people who have had major surgeries in states R1023, R1019, R1016, and R1014 are also relatively high. This suggests that these states may also face challenges related to high healthcare costs for people who have had major surgeries.
- The range of average hospitalization costs for people who have had major surgeries across different states is large. This suggests that there is a significant variation in the cost of healthcare for people who have had major surgeries, depending on where they live.

Healthcare organizations can use this information to make informed decisions about resource allocation and cost-effective strategies. For example, healthcare organizations may want to consider investing in more resources in states with high average hospitalization costs for people who have had major surgeries. They may also want to work with providers in these states to identify ways to reduce the cost of healthcare.

Insights and conclusions:

- State R1011 has the highest average hospitalization cost, while state R1025 has the lowest. This information can guide resource allocation and cost-effective strategies, especially in regions with higher healthcare expenses.
- Tier 2 hospitals have the most diabetic customers. This is due to the number of factors, including the location of Tier 2 hospitals, their affordability, and their convenience.
- The average BMI of customers with major surgeries is approximately 31.01. This suggests that individuals who have had major surgeries tend to have a higher average BMI, which could be significant for healthcare providers in understanding the health profiles of this patient group.
- Tier-1 cities have the highest number of customers, followed by tier-2 and tier-3 cities. This distribution of customers across different city tiers provides healthcare organizations with essential insights into where their services are in demand and how they can better cater to the specific needs of each region.
- Healthcare organizations can use the information about the distribution of healthcare costs and other health-related metrics across different regions and populations to make better decisions about how to allocate resources, develop programs and services, and improve the quality and affordability of healthcare for all.
- For example, healthcare organizations may want to consider investing more resources in state R1011, where healthcare costs are highest. They may also want to work with providers in state R1011 to identify ways to reduce the cost of healthcare. Additionally, healthcare organizations can develop cost-effective strategies for providing healthcare services in state R1011, such as telemedicine or other remote healthcare options.
- To better serve the needs of diabetic patients, healthcare organizations may want to consider partnering with Tier 2 hospitals to improve access to care or develop programs that support diabetic patients in managing their condition.
- Healthcare organizations may also want to develop programs and services that are tailored to the specific needs of patients who have had major surgeries, such as weight management programs or support groups.
- Finally, healthcare organizations can use the information about the distribution of customers across different city tiers to plan for future growth and development, such as identifying areas where they need to build new facilities or expand existing ones.

Thank you!