UNCOVERING HEALTHCARE PATTERNS AND PATIENT INSIGHTS THROUGH MYSQL DATA ANALYSIS

MARK ANTHONY A. BUNA
Data Analyst

🧠 PROBLEM STATEMENT

- Healthcare organizations handle large volumes of patient data, yet actionable insights are often hidden in scattered records. Without efficient analysis, trends in admissions, diagnoses, treatments, and billing go unnoticed. This limits the potential for improving operations and delivering better patient outcomes. This project uses MySQL to uncover key healthcare patterns by exploring patient demographics, hospital performance, doctor workloads, and insurance billing data.

🎯 OBJECTIVE

- To utilize MySQL for analyzing healthcare data, with the goal of identifying patient trends, optimizing hospital operations, and enhancing data-driven insights.
- Uncover insights into healthcare trends by tackling these analytical questions:
    - ***Basic Queries (Data Exploration)***
        1. *How many unique patients are recorded in the dataset?*
        2. *Retrieve a list of distinct hospitals where patients were admitted.*
        3. *Find all patients diagnosed with Diabetes, displaying their name, age, and hospital.*
        4. *Count the number of male and female patients in the dataset.*
        5. *What are the different types of admissions, and how many patients fall under each category?*
    - ***Aggregation & Grouping***
        6. *Count the number of patients admitted per hospital, ordered by highest admissions.*
        7. *What is the average billing amount for each admission type?*

8. *Identify the hospital with the highest total billing amount across all admissions.*
9. *Retrieve the top 5 most common medical conditions, sorted by number of patients diagnosed.*
10. *Which doctor has treated the most patients, and how many have they treated?*

- ***Date & Time Analysis***
  11. *What is the earliest and latest admission date recorded?*
  12. *Count the number of patients admitted in the last 30 days from the most recent admission date.*
  13. *Calculate the average length of stay for patients (difference between discharge and admission date).*
  14. *Find the month and year with the highest number of admissions and the corresponding count.*
  15. *Count how many emergency admissions happened on weekends.*
- ***Advanced Queries (Joins, Subqueries, and Window Functions)***
  16. *Identify the most prescribed medication, along with the number of times it was prescribed.*
  17. *Retrieve details of patients with "Abnormal" test results, including their medical condition and assigned doctor.*
  18. *Which insurance provider has covered the highest total billing amount?*
  19. *Identify patients prescribed both "Aspirin" and "Ibuprofen" during their admission.*
  20. *Find the room number with the highest patient occupancy and how many patients stayed there.*

📦 FINAL PORTFOLIO DELIVERABLES

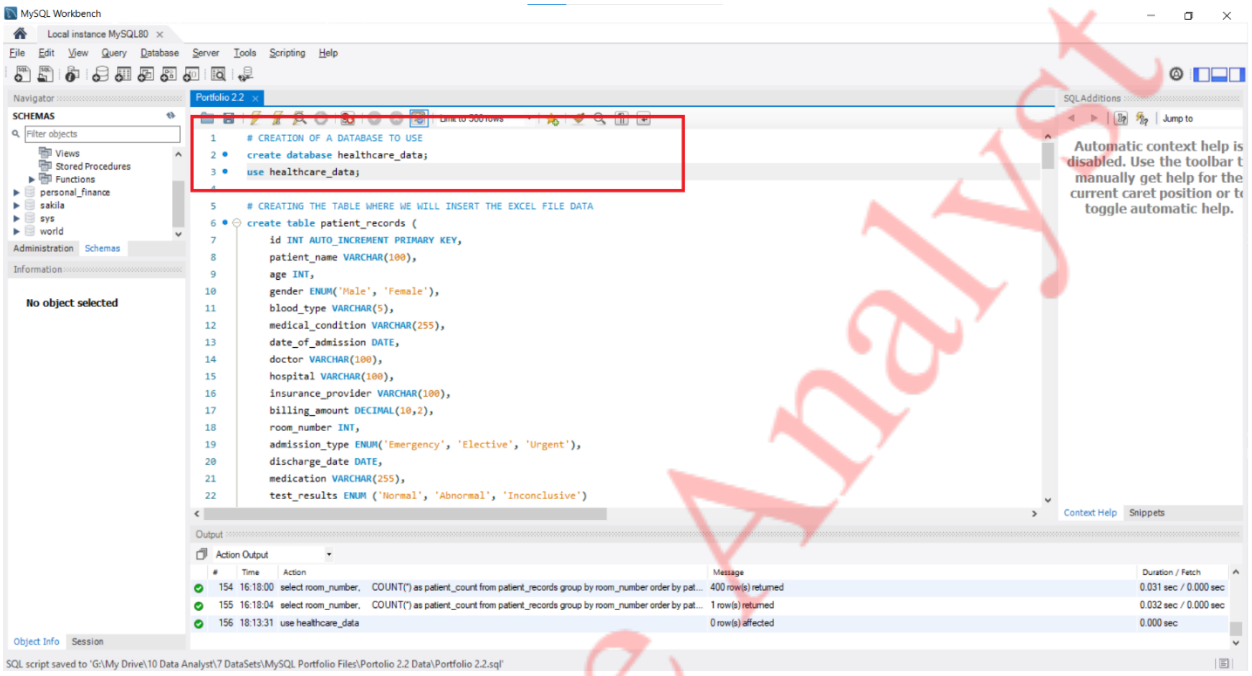
✅ Built a MySQL database to analyze healthcare trends and billing, developed SQL queries to explore patient data and uncover insights, and documented the entire project in a GitHub repository.

🔗 LINKS

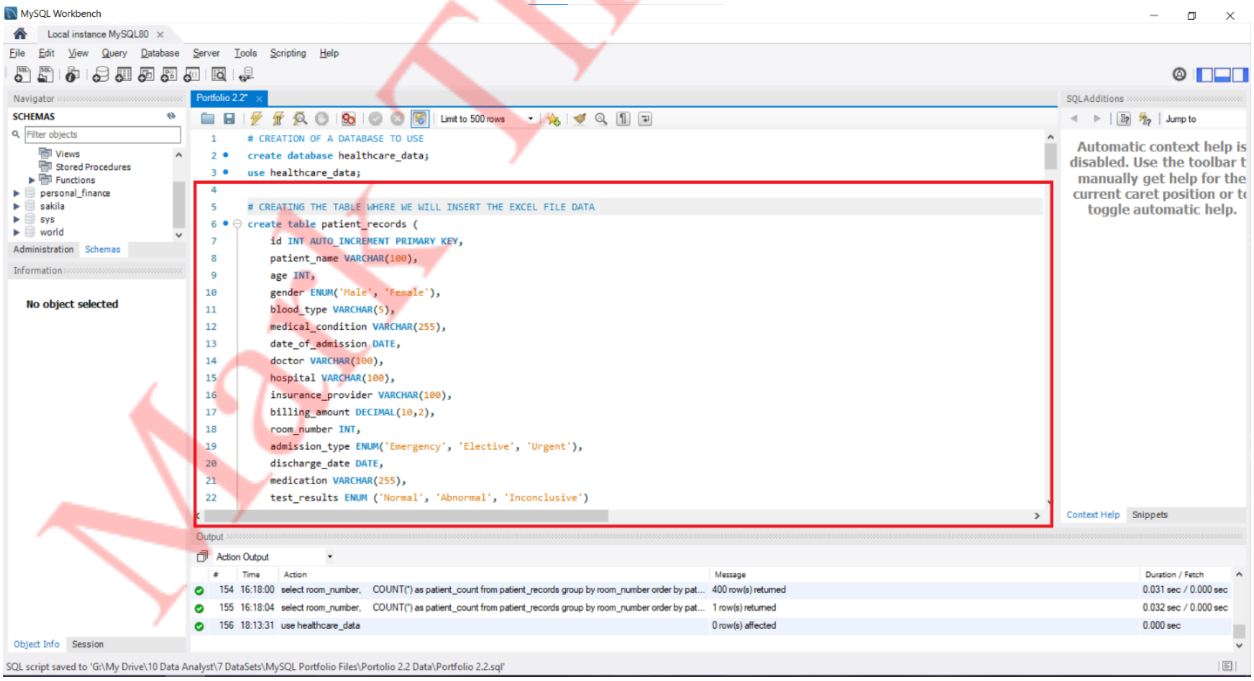
- Dataset*:* *https://www.kaggle.com/datasets/parthgosavi/healthcare-dataset*
- Github*:* *https://github.com/marktheanalyst103/mysql-scripts/blob/main/Uncovering%20Healthcare%20Patterns%20and%20Patient%20Insights%20through%20Mysql%20Data%20Analysis.sql*

PROJECT STEPS

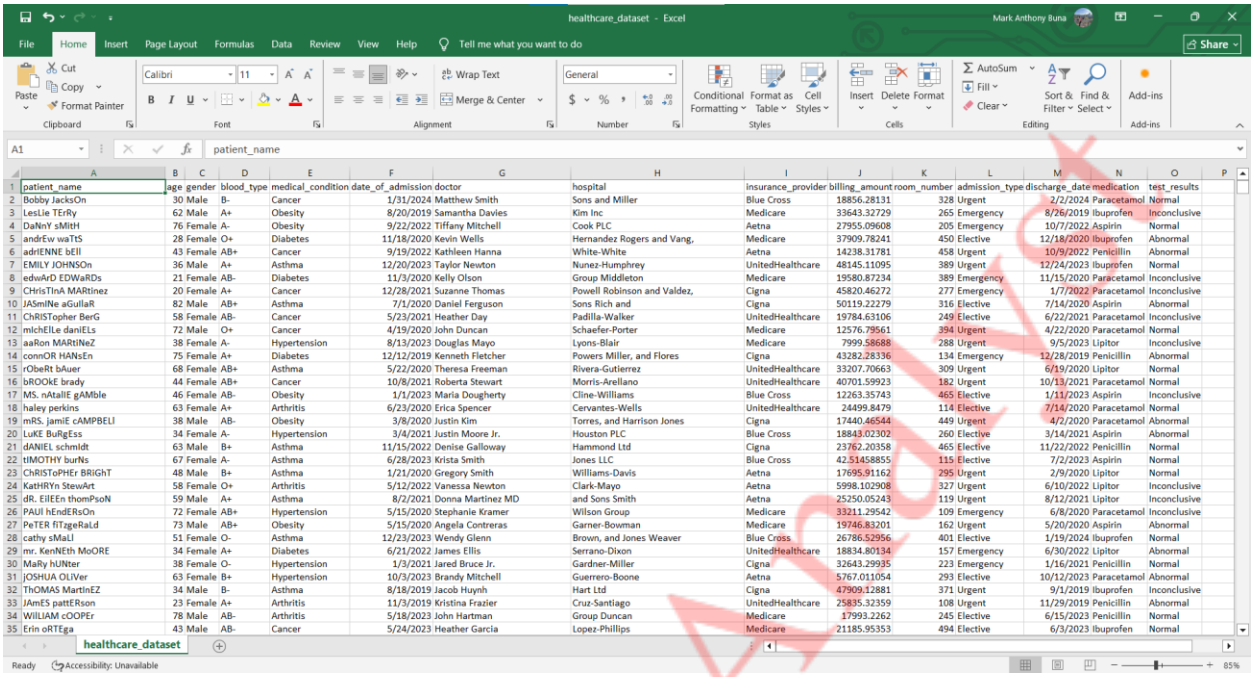
I. SET UP MySQL DATABASE



II. PREPARATION OF THE MYSQL TABLE FOR INSERTING EXCEL FILE DATA

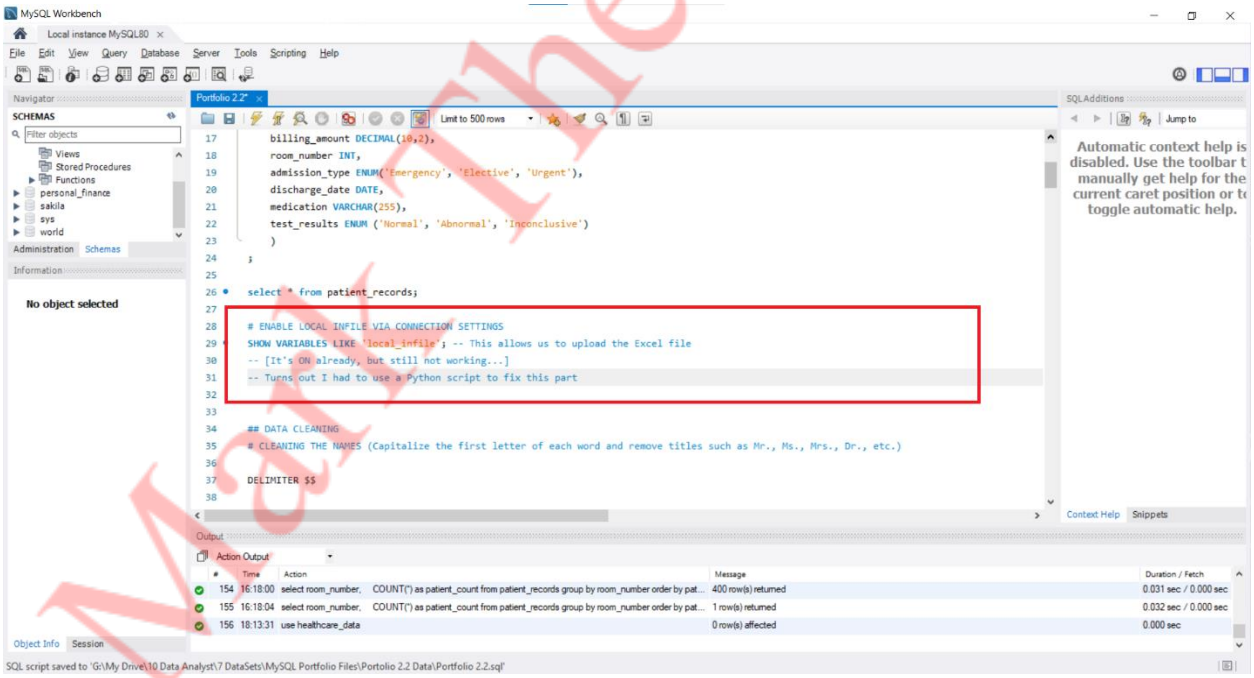


III. INSERTION OF EXCEL FILE DATA



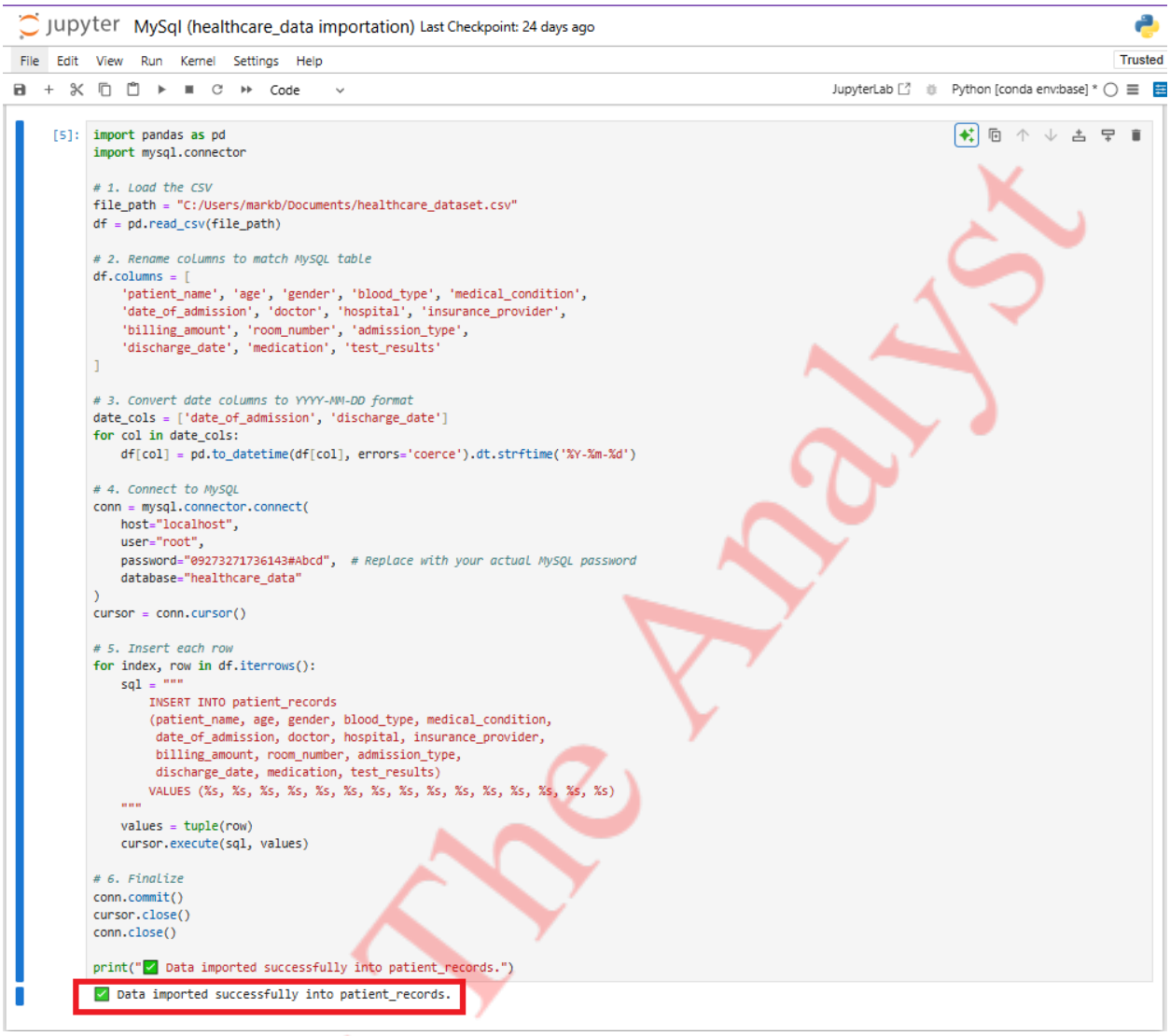
patient_name	age	gender	blood_type	medical_condition	date_of_admission	doctor	hospital	insurance_provider	billing_amount	room_number	admission_type	discharge_date	medication	test_results
Bobby JacksOn	30	Male	B-	Cancer	1/31/2024	Matthew Smith	Sons and Miller	Blue Cross	18856.28131	328	Urgent	2/2/2024	Paracetamol	Normal
LesLie TErRy	62	Male	A+	Obesity	8/20/2019	Samantha Davies	Kim Inc	Medicare	33643.32729	265	Emergency	8/26/2019	Ibuprofen	Inconclusive
DaNnY sMitH	76	Female	A-	Obesity	9/22/2022	Tiffany Mitchell	Cook PLC	Aetna	27955.09608	205	Emergency	10/7/2022	Aspirin	Normal
andrEw waTtS	28	Female	O+	Diabetes	11/18/2020	Kevin Wells	Hernandez Rogers and Vang,	Medicare	37909.78241	450	Elective	12/18/2020	Ibuprofen	Abnormal
adrIENNE bEll	43	Female	AB+	Cancer	9/19/2022	Kathleen Hanna	White-White	Aetna	14238.31781	458	Urgent	10/9/2022	Penicillin	Abnormal
EMILY JOHNSOn	36	Male	A+	Asthma	12/20/2023	Taylor Newton	Nunez-Humphrey	UnitedHealthcare	48145.11095	389	Urgent	12/24/2023	Ibuprofen	Normal
edwArD EDWaRDs	21	Female	AB-	Diabetes	11/3/2020	Kelly Olson	Group Middleton	Medicare	19580.87234	389	Emergency	11/15/2020	Paracetamol	Inconclusive
CHrisTInA MARtinez	20	Female	A+	Cancer	12/28/2021	Suzanne Thomas	Powell Robinson and Valdez,	Cigna	45820.46272	277	Emergency	1/7/2022	Paracetamol	Inconclusive
JASmINe aGullaR	82	Male	AB+	Asthma	7/1/2020	Daniel Ferguson	Sons Rich and	Cigna	50119.22279	316	Elective	7/14/2020	Aspirin	Abnormal
ChRISTopher BerG	58	Female	AB-	Cancer	5/23/2021	Heather Day	Padilla-Walker	UnitedHealthcare	19784.63106	249	Elective	6/22/2021	Paracetamol	Inconclusive
mIchElLe danIELs	72	Male	O+	Cancer	4/19/2020	John Duncan	Schaefer-Porter	Medicare	12576.79561	394	Urgent	4/22/2020	Paracetamol	Normal
aaRon MARtINeZ	38	Female	A-	Hypertension	8/13/2023	Douglas Mayo	Lyons-Blair	Medicare	7999.58688	288	Urgent	9/5/2023	Lipitor	Inconclusive
connOR HANsEn	75	Female	A+	Diabetes	12/12/2019	Kenneth Fletcher	Powers Miller, and Flores	Cigna	43282.28336	134	Emergency	12/28/2019	Penicillin	Abnormal
rObeRt bAuer	68	Female	AB+	Asthma	5/22/2020	Theresa Freeman	Rivera-Gutierrez	UnitedHealthcare	33207.70663	309	Urgent	6/19/2020	Lipitor	Normal
bROOkE brady	44	Female	AB+	Cancer	10/8/2021	Roberta Stewart	Morris-Arellano	UnitedHealthcare	40701.59923	182	Urgent	10/13/2021	Paracetamol	Normal
MS. nAtalIE gAMble	46	Female	AB-	Obesity	1/1/2023	Maria Dougherty	Cline-Williams	Blue Cross	12263.35743	465	Elective	1/11/2023	Aspirin	Inconclusive
haley perkins	63	Female	A+	Arthritis	6/23/2020	Erica Spencer	Cervantes-Wells	UnitedHealthcare	24499.8479	114	Elective	7/14/2020	Paracetamol	Normal
mRS. jamIE cAMPBELl	38	Male	AB-	Obesity	3/8/2020	Justin Kim	Torres, and Harrison Jones	Cigna	17440.46544	449	Urgent	4/2/2020	Paracetamol	Abnormal
LuKE BuRgEss	34	Female	A-	Hypertension	3/4/2021	Justin Moore Jr.	Houston PLC	Blue Cross	18843.02302	260	Elective	3/14/2021	Aspirin	Abnormal
dANIEL schmIdt	63	Male	B+	Asthma	11/15/2022	Denise Galloway	Hammond Ltd	Cigna	23762.20358	465	Elective	11/22/2022	Penicillin	Normal
tIMOTHY burNs	67	Female	A-	Asthma	6/28/2023	Krista Smith	Jones LLC	Blue Cross	42.51458855	115	Elective	7/2/2023	Aspirin	Normal
ChRISTOPHEr BRIGhT	48	Male	B+	Asthma	1/21/2020	Gregory Smith	Williams-Davis	Aetna	17695.91162	295	Urgent	2/9/2020	Lipitor	Normal
KatHRYn StewArt	58	Female	O+	Arthritis	5/12/2022	Vanessa Newton	Clark-Mayo	Aetna	5998.102908	327	Urgent	6/10/2022	Lipitor	Inconclusive
dR. EllEEn thomPsoN	59	Male	A+	Asthma	8/2/2021	Donna Martinez MD	and Sons Smith	Aetna	25250.05243	119	Urgent	8/12/2021	Lipitor	Inconclusive
PAUl hEndERsOn	72	Female	AB+	Hypertension	5/15/2020	Stephanie Kramer	Wilson Group	Medicare	33211.29542	109	Emergency	6/8/2020	Paracetamol	Inconclusive
PeTER fITzgeRaLd	73	Male	AB+	Obesity	5/15/2020	Angela Contreras	Garner-Bowman	Medicare	19746.83201	162	Urgent	5/20/2020	Aspirin	Abnormal
cathy sMaLl	51	Female	O-	Asthma	12/23/2023	Wendy Glenn	Brown, and Jones Weaver	Blue Cross	26786.52956	401	Elective	1/19/2024	Ibuprofen	Normal
mr. KenNEth MoORE	34	Female	A+	Diabetes	6/21/2022	James Ellis	Serrano-Dixon	UnitedHealthcare	18834.80134	157	Emergency	6/30/2022	Lipitor	Abnormal
MaRy hUNter	38	Female	O-	Hypertension	1/3/2021	Jared Bruce Jr.	Gardner-Miller	Cigna	32643.29935	223	Emergency	1/16/2021	Penicillin	Normal
jOSHUA OLiVer	63	Female	B+	Hypertension	10/3/2023	Brandy Mitchell	Guerrero-Boone	Aetna	5767.011054	293	Elective	10/12/2023	Paracetamol	Abnormal
ThOMAS MartInEZ	34	Male	B-	Asthma	8/18/2019	Jacob Huynh	Hart Ltd	Cigna	47909.12881	371	Urgent	9/1/2019	Ibuprofen	Inconclusive
JAmES pattERson	23	Female	A+	Arthritis	11/3/2019	Kristina Frazier	Cruz-Santiago	UnitedHealthcare	25835.32359	108	Urgent	11/29/2019	Penicillin	Abnormal
WiLLIAM cOOPEr	78	Male	AB-	Arthritis	5/18/2023	John Hartman	Group Duncan	Medicare	17993.2262	245	Elective	6/15/2023	Penicillin	Normal
Erin oRTega	43	Male	AB-	Cancer	5/24/2023	Heather Garcia	Lopez-Phillips	Medicare	21185.95353	494	Elective	6/3/2023	Ibuprofen	Normal

- *The data from the Microsoft Excel file that I need to upload and use for analysis in MySQL*



```
billing_amount DECIMAL(10,2),
room_number INT,
admission_type ENUM('Emergency', 'Elective', 'Urgent'),
discharge_date DATE,
medication VARCHAR(255),
test_results ENUM ('Normal', 'Abnormal', 'Inconclusive')
)
;

select * from patient_records;

# ENABLE LOCAL INFILE VIA CONNECTION SETTINGS
SHOW VARIABLES LIKE 'local_infile'; -- This allows us to upload the Excel file
-- [It's ON already, but still not working...]
-- Turns out I had to use a Python script to fix this part


## DATA CLEANING
# CLEANING THE NAMES (Capitalize the first letter of each word and remove titles such as Mr., Ms., Mrs., Dr., etc.)

DELIMITER $$
```

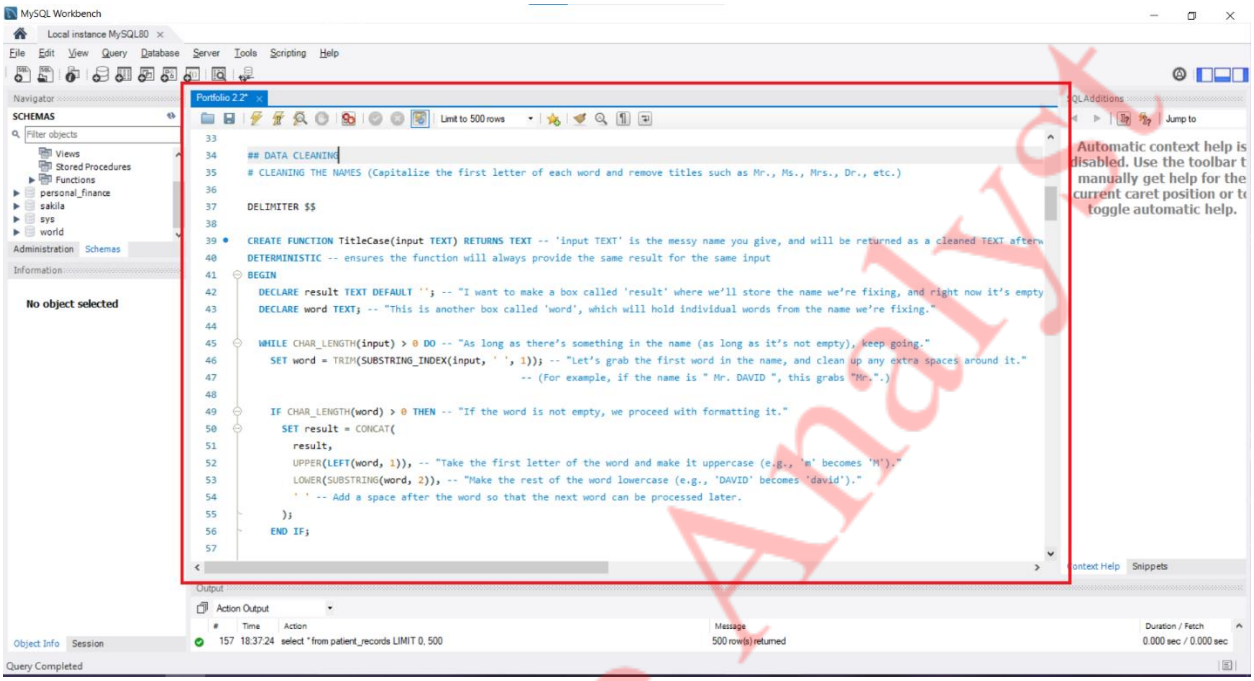
- *For certain reasons, MySQL was not allowing me to upload the excel file so I had to find ways how to do it, and using python script is one that worked for me.*



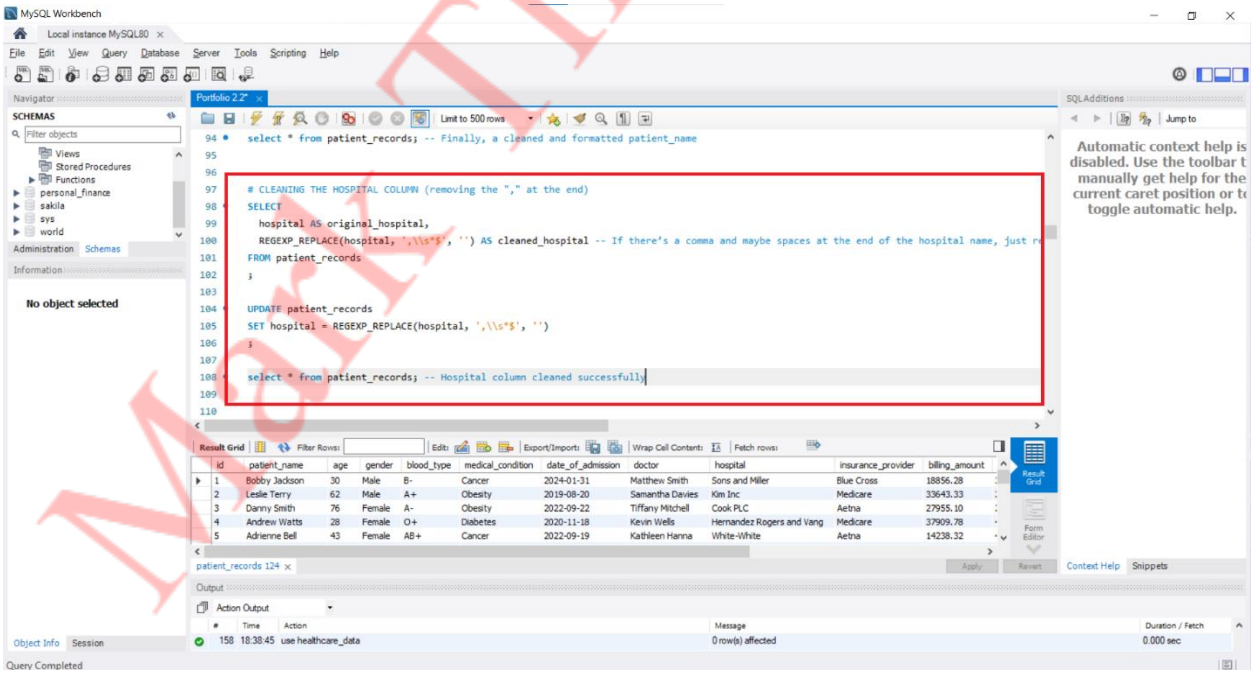
- *The Python script I used to upload the Excel file to MySQL for later analysis.*

IV. DATA CLEANING

- Cleaning the names (Capitalize the first letter of each word and remove titles such as Mr., Ms., Mrs., Dr., etc.)



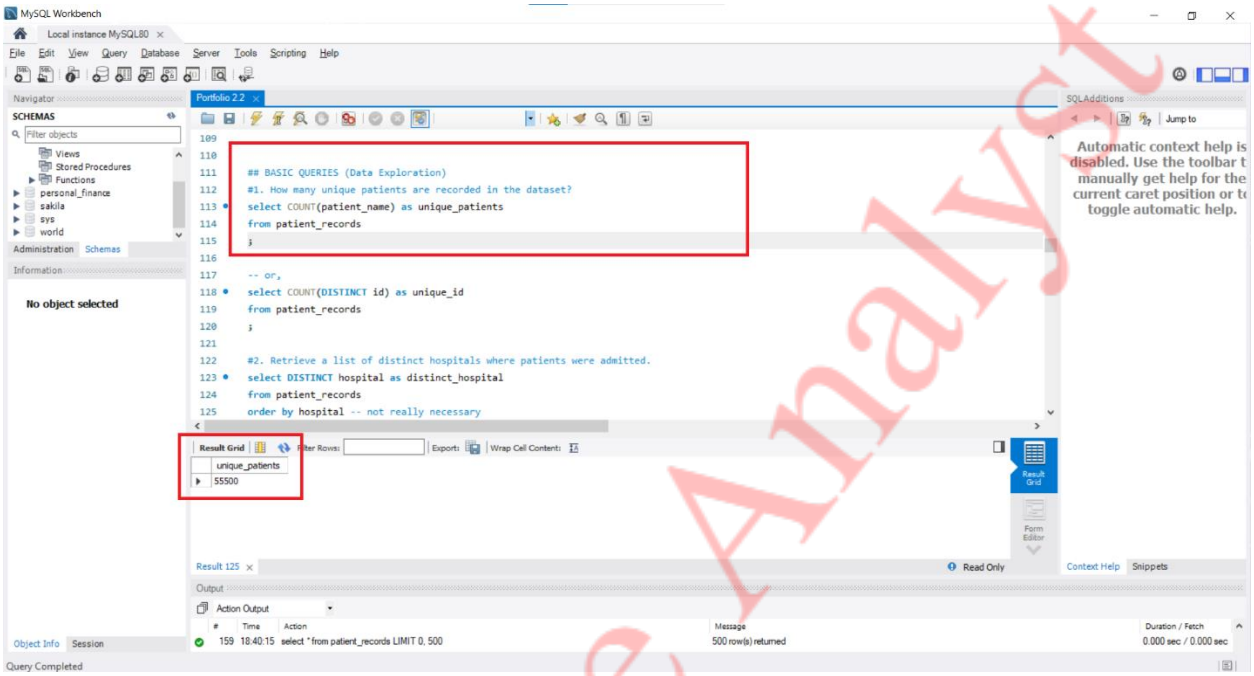
- Cleaning the Hospital Column (removing the "," at the end)



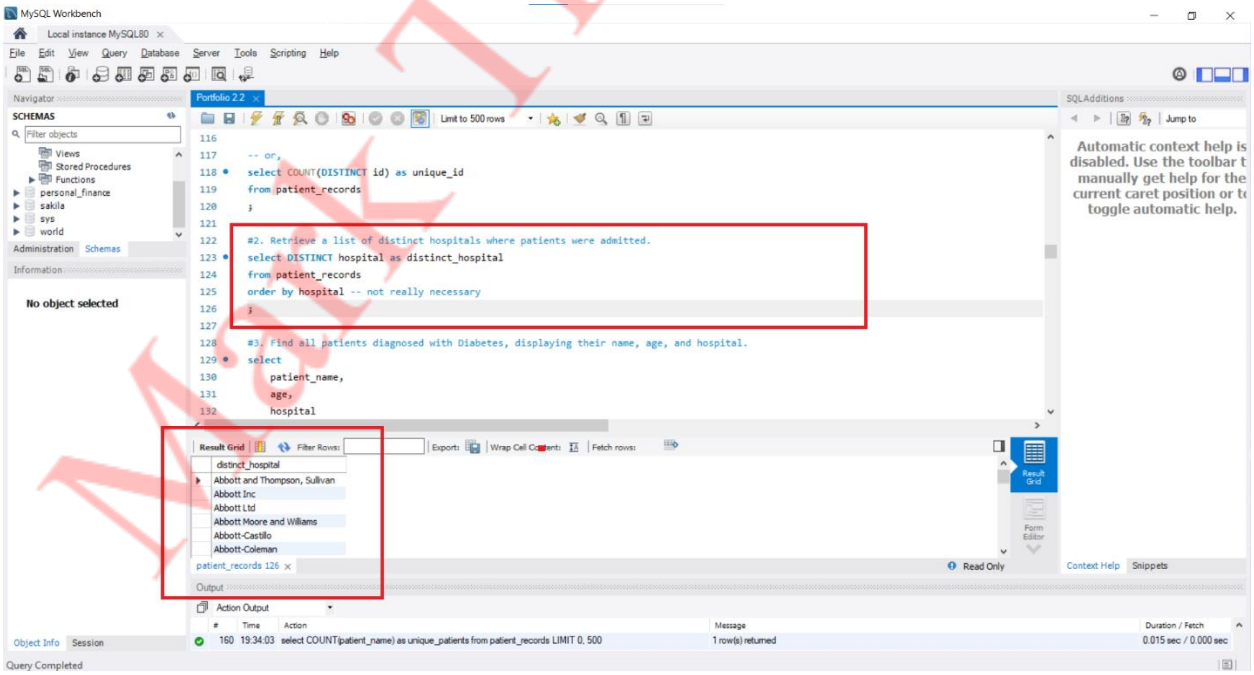
V. MySQL QUERIES

Basic Queries (Data Exploration)

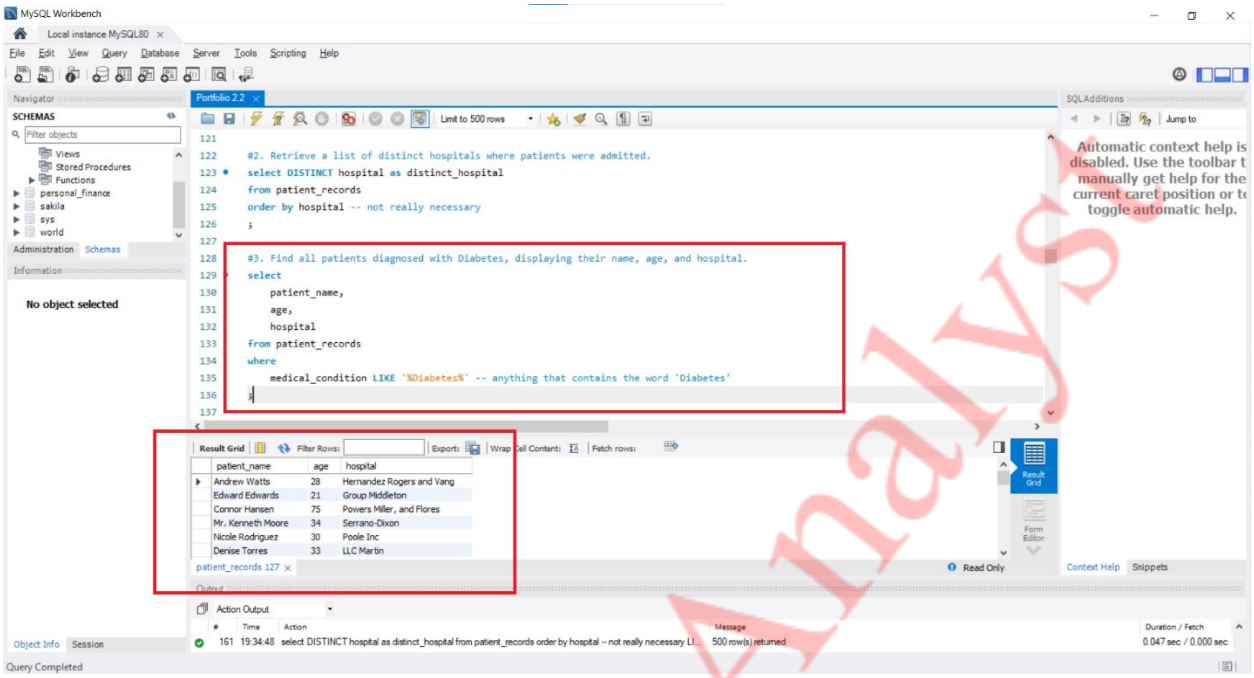
1. How many unique patients are recorded in the dataset?



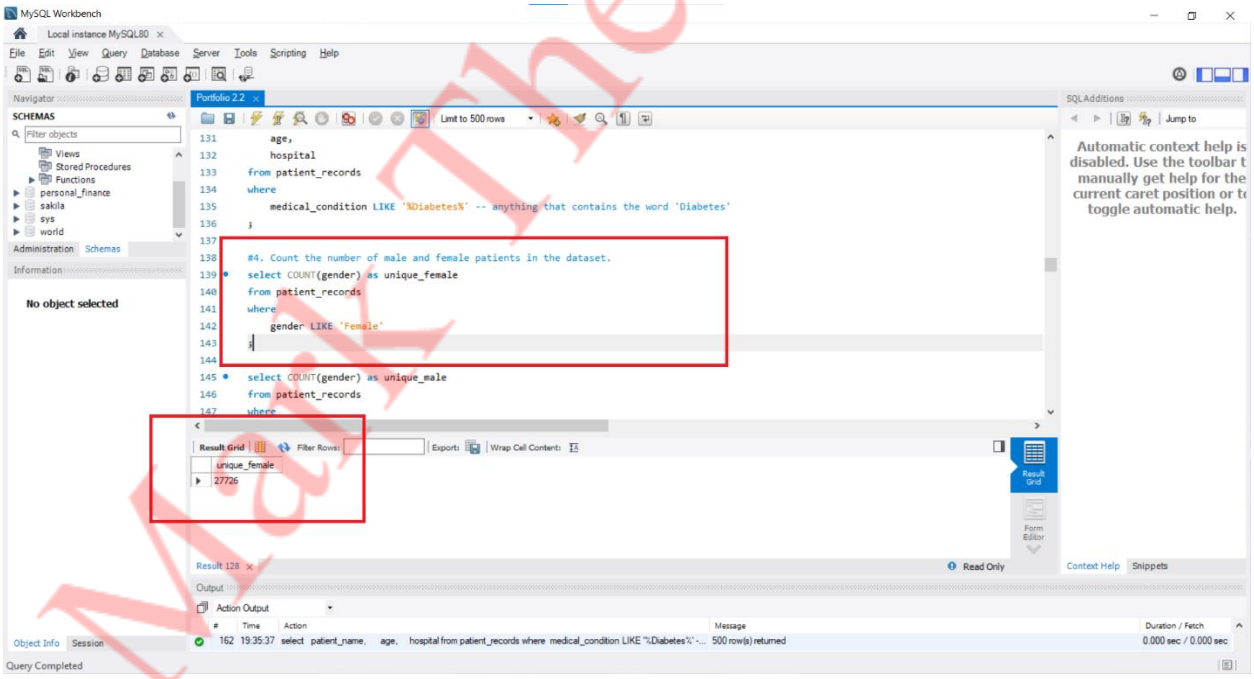
2. Retrieve a list of distinct hospitals where patients were admitted.

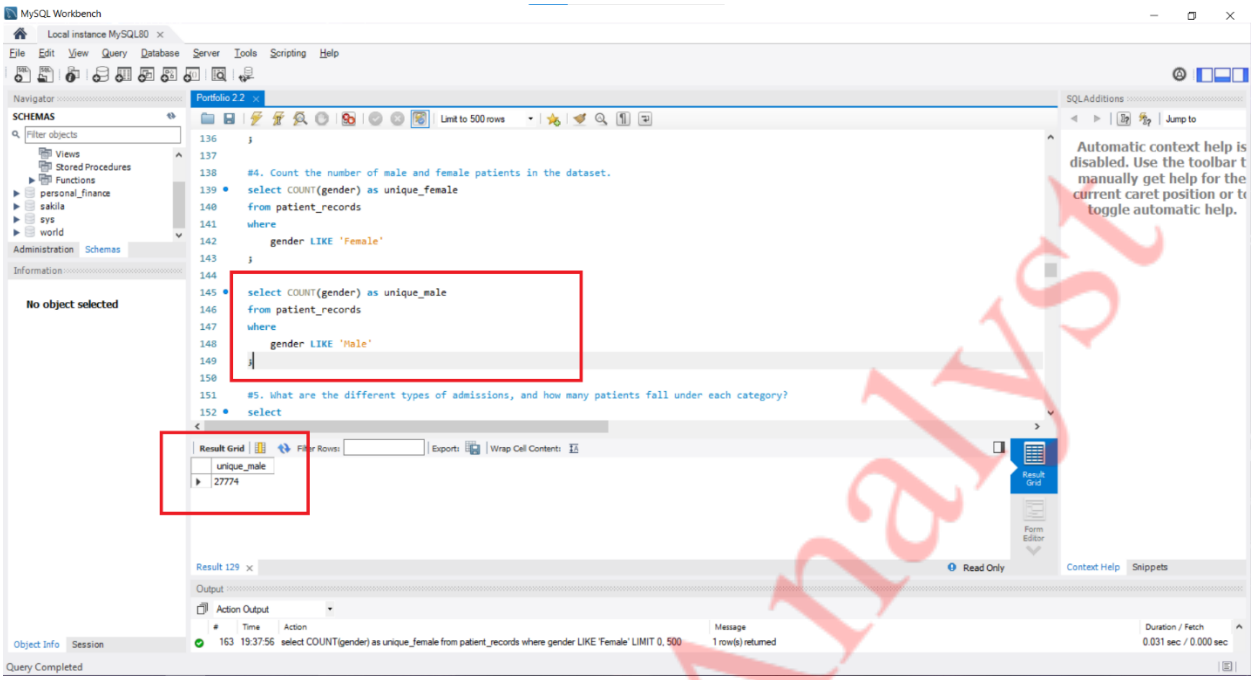


3. Find all patients diagnosed with **Diabetes**, displaying their name, age, and hospital.

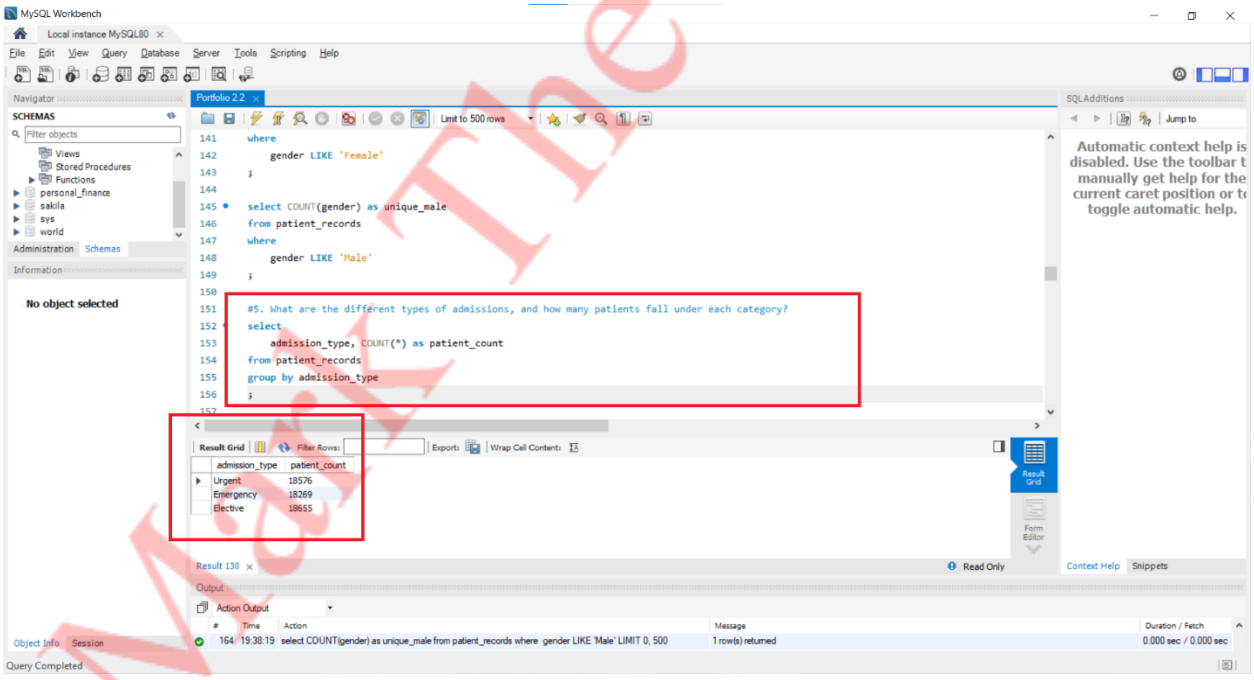


4. Count the number of **male** and **female** patients in the dataset.



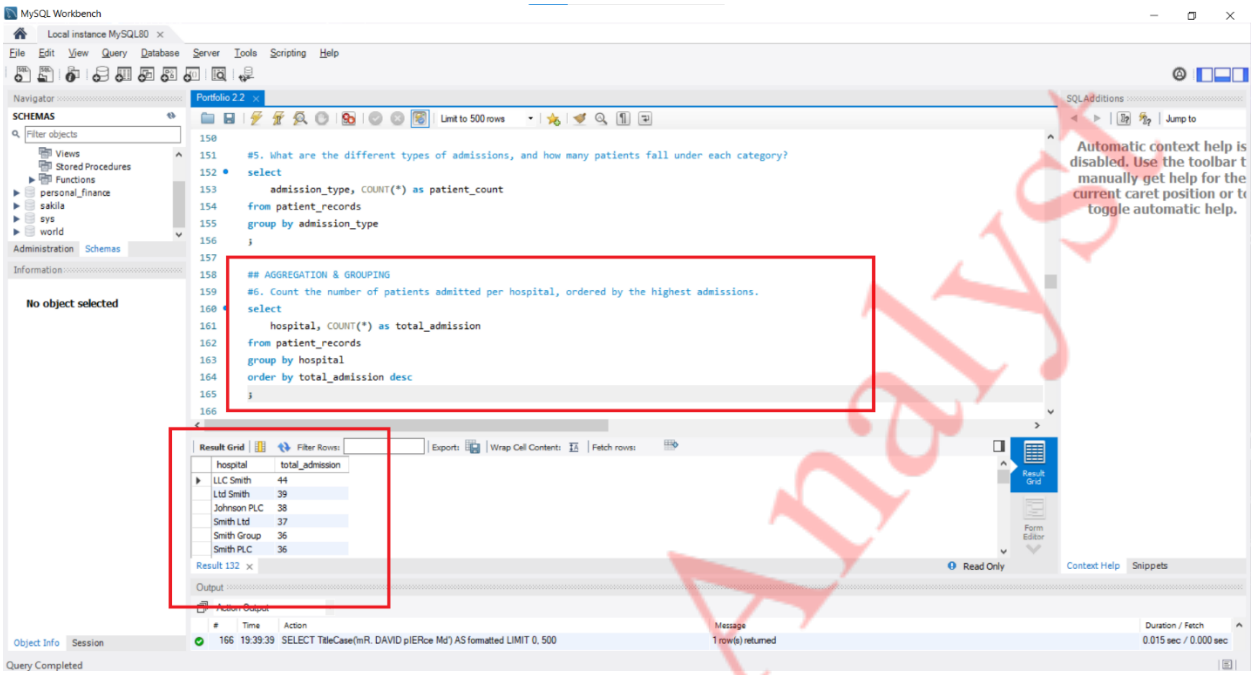


5. What are the different types of admissions, and how many patients fall under each category?

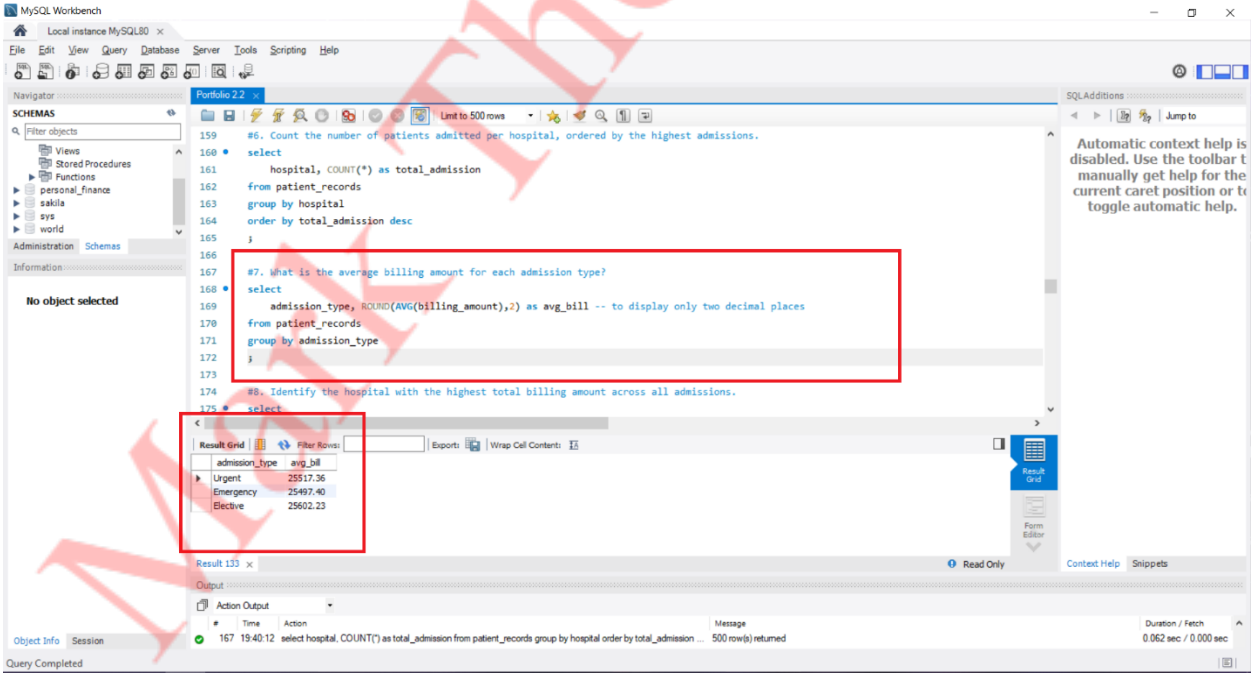


Aggregation & Grouping

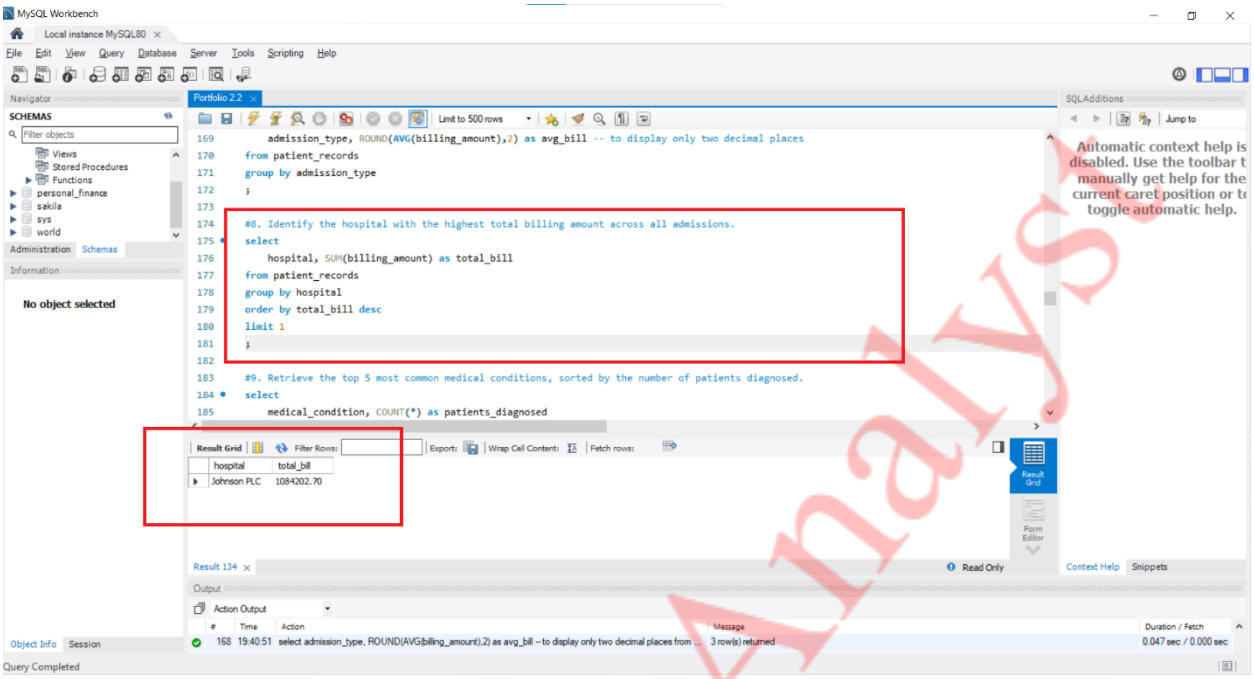
6. Count the number of **patients admitted per hospital**, ordered by the highest admissions.



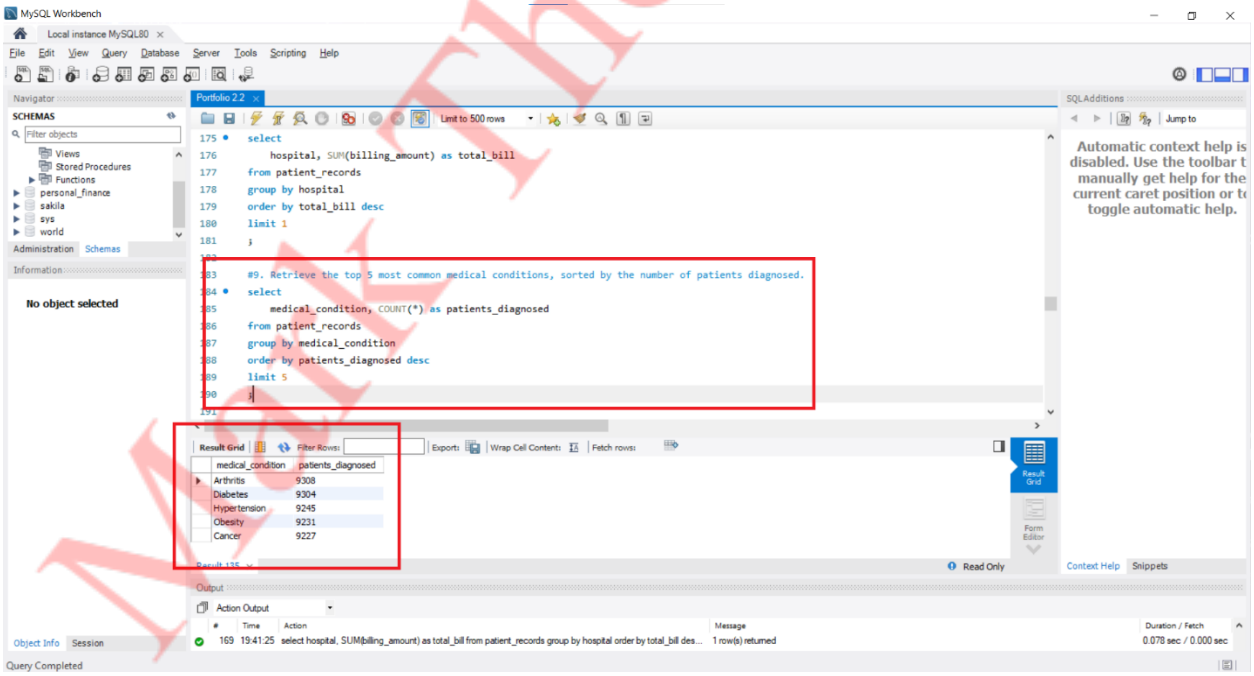
7. What is the **average billing amount** for each admission type?



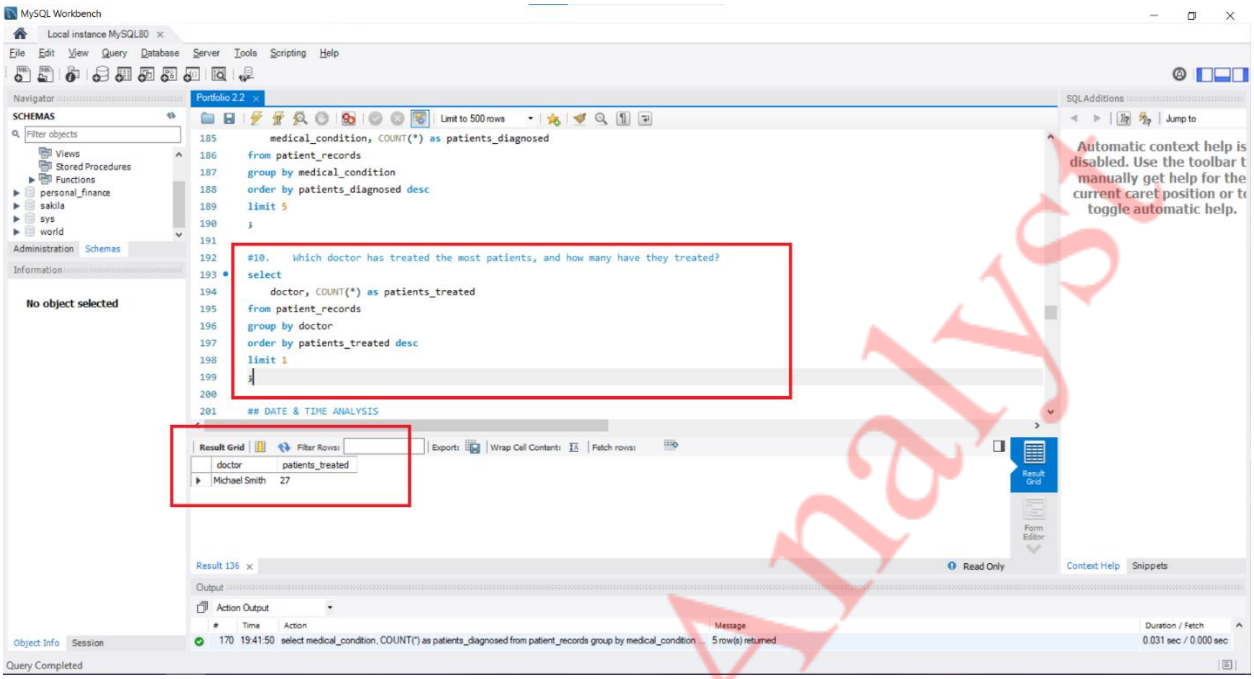
8. Identify the **hospital with the highest total billing amount** across all admissions.



9. Retrieve the **top 5 most common medical conditions**, sorted by the number of patients diagnosed.

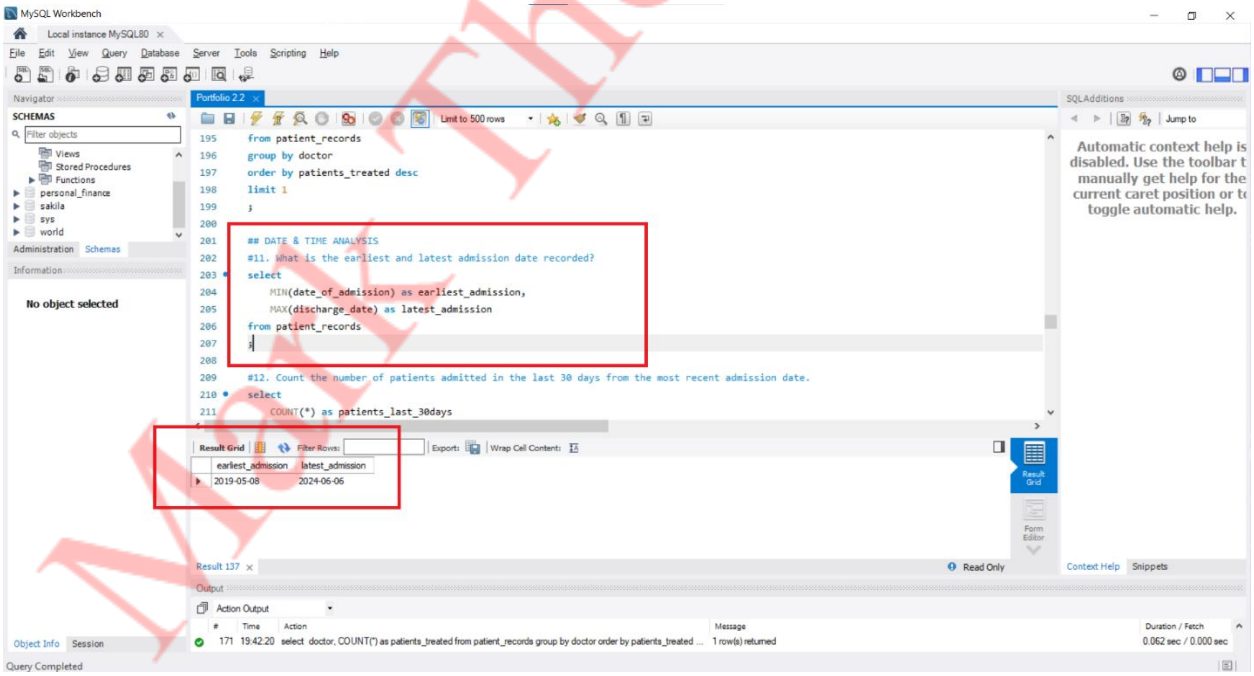


10. Which doctor has treated **the most patients**, and how many have they treated?

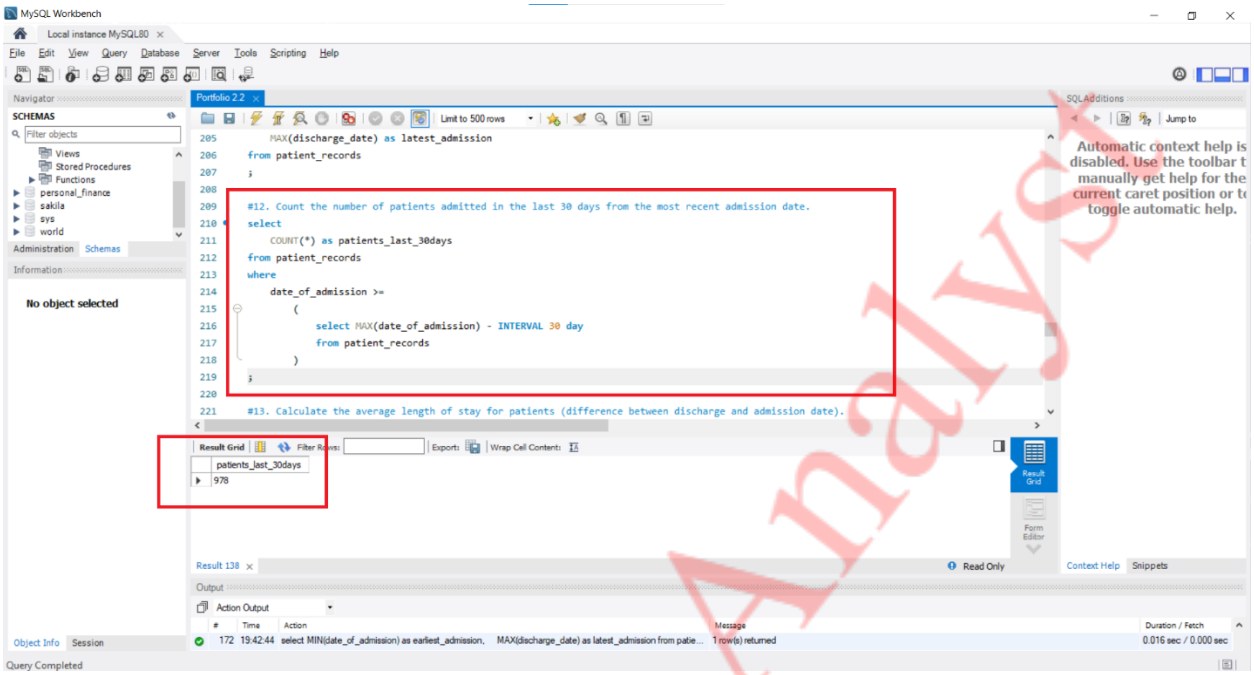


Date & Time Analysis

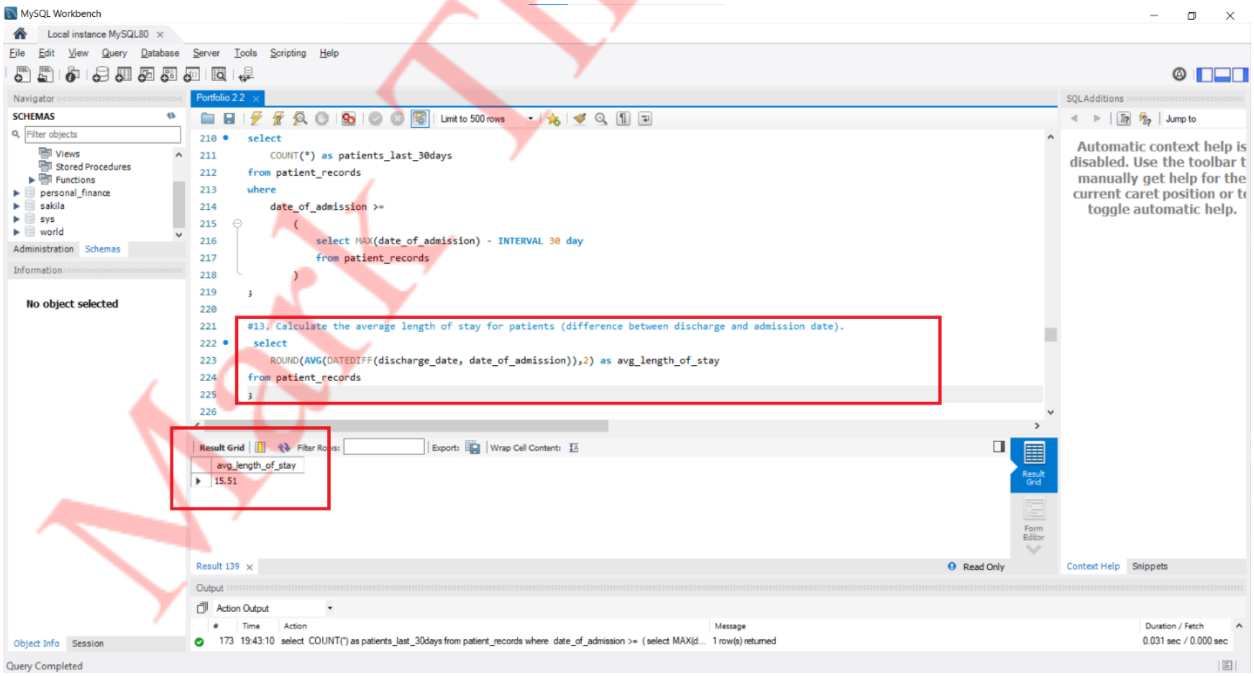
11. What is the **earliest and latest admission date** recorded?



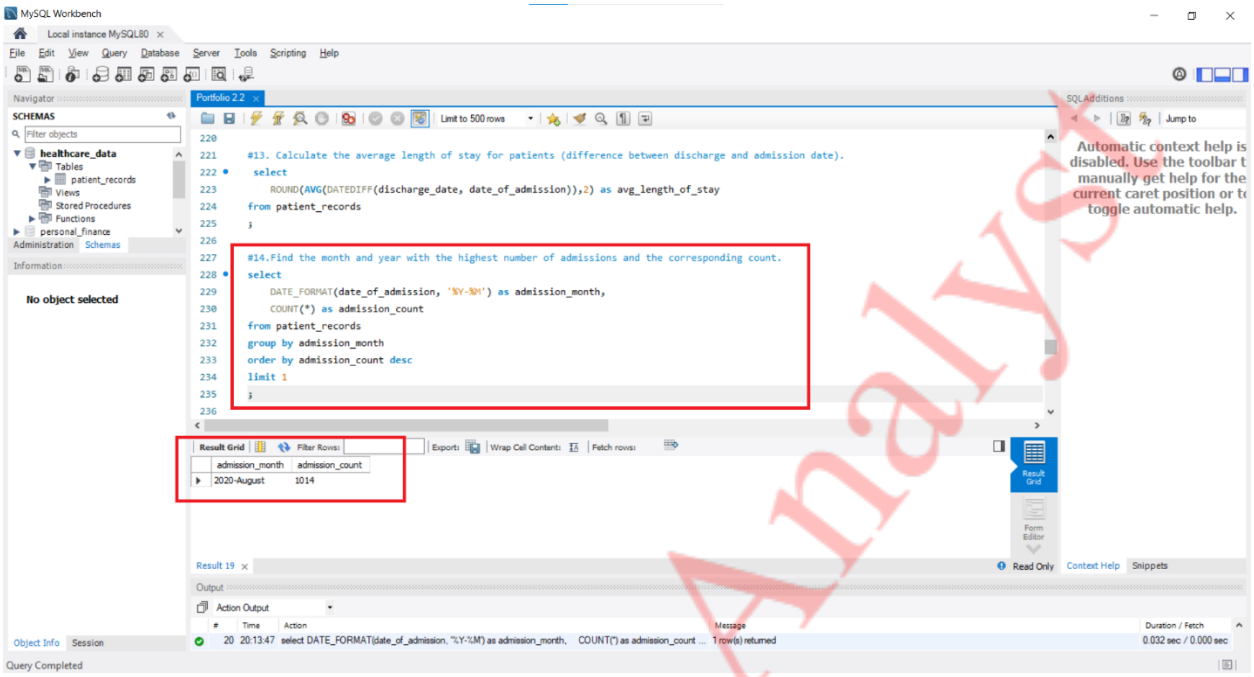
12. Count the number of **patients admitted in the last 30 days** from the most recent admission date.



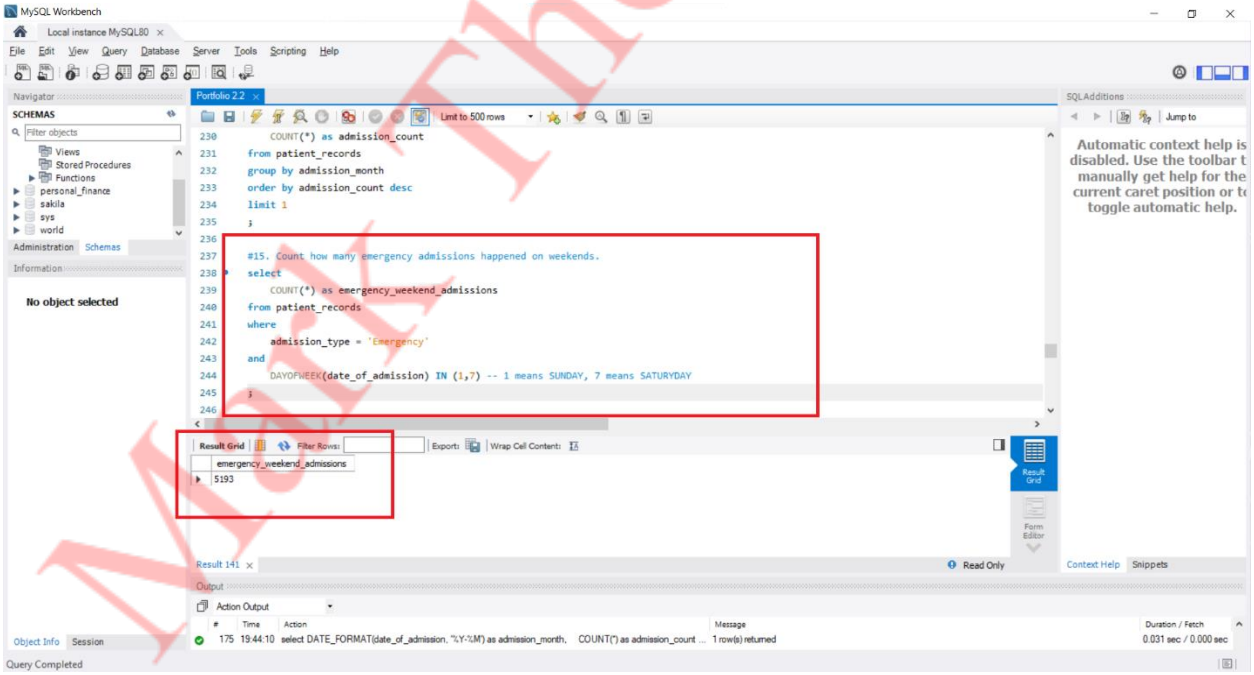
13. Calculate the **average length of stay** for patients (difference between discharge and admission date).



14. Find the **month and year with the highest number of admissions** and the corresponding count.

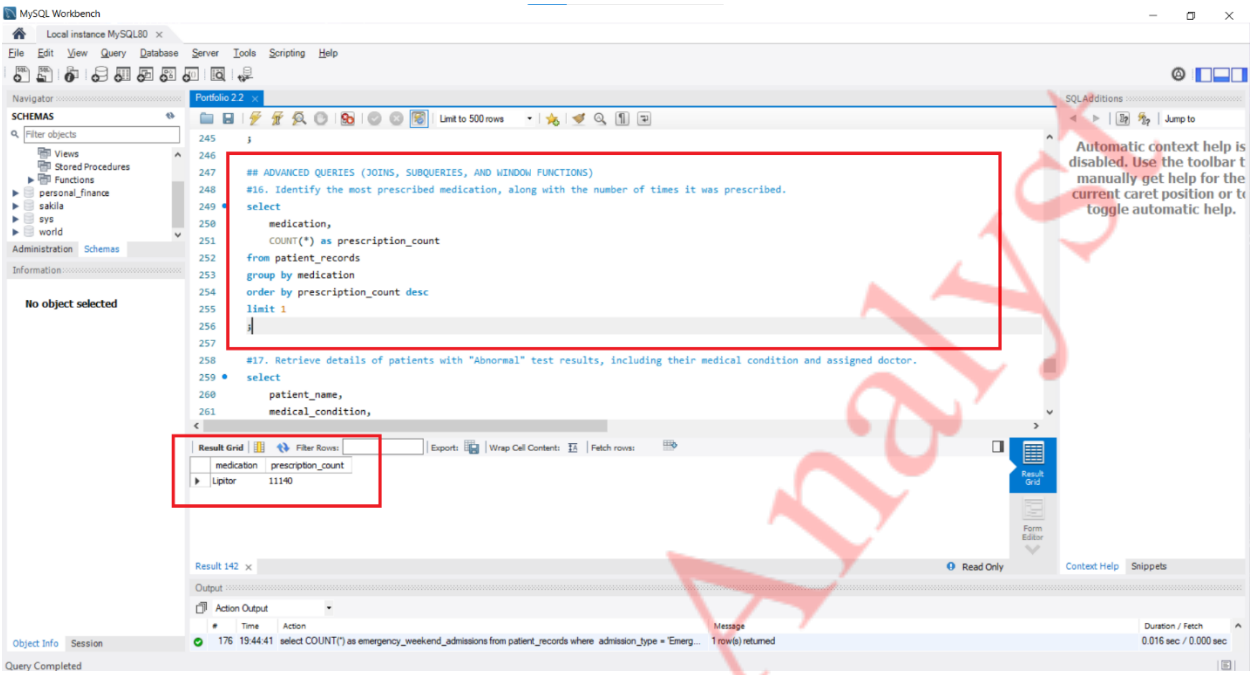


15. Count how many **emergency admissions happened on weekends**.

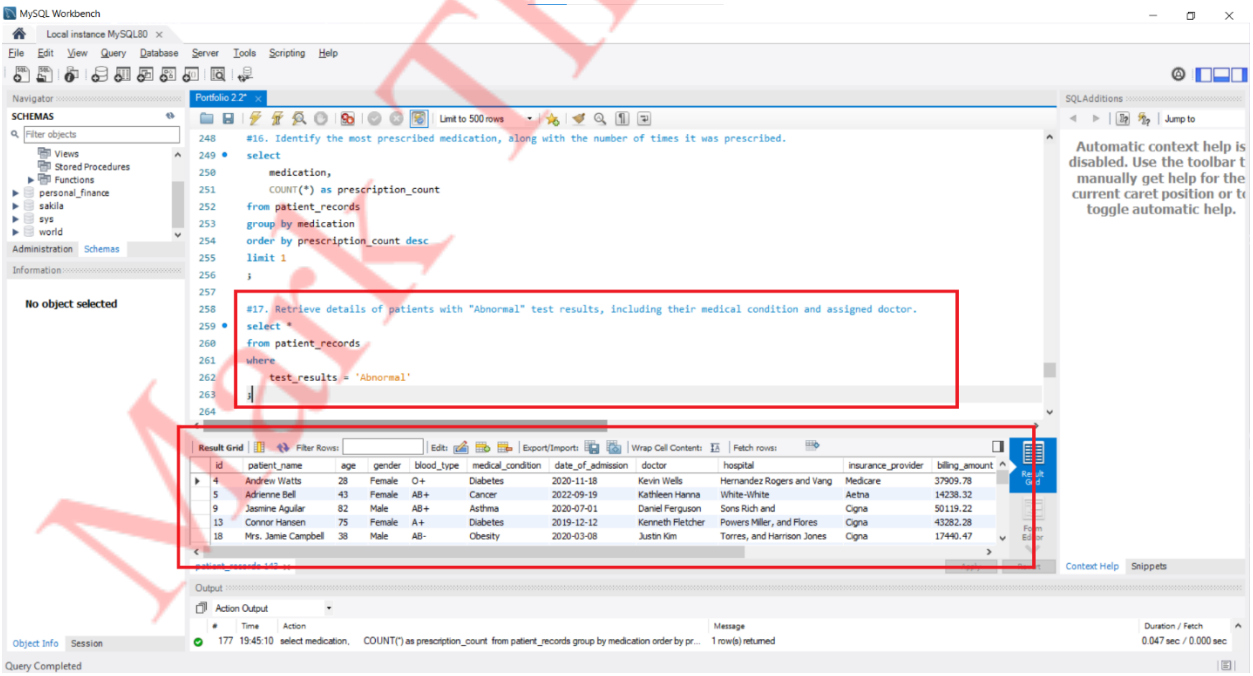


Advanced Queries (Joins, Subqueries, and Window Functions)

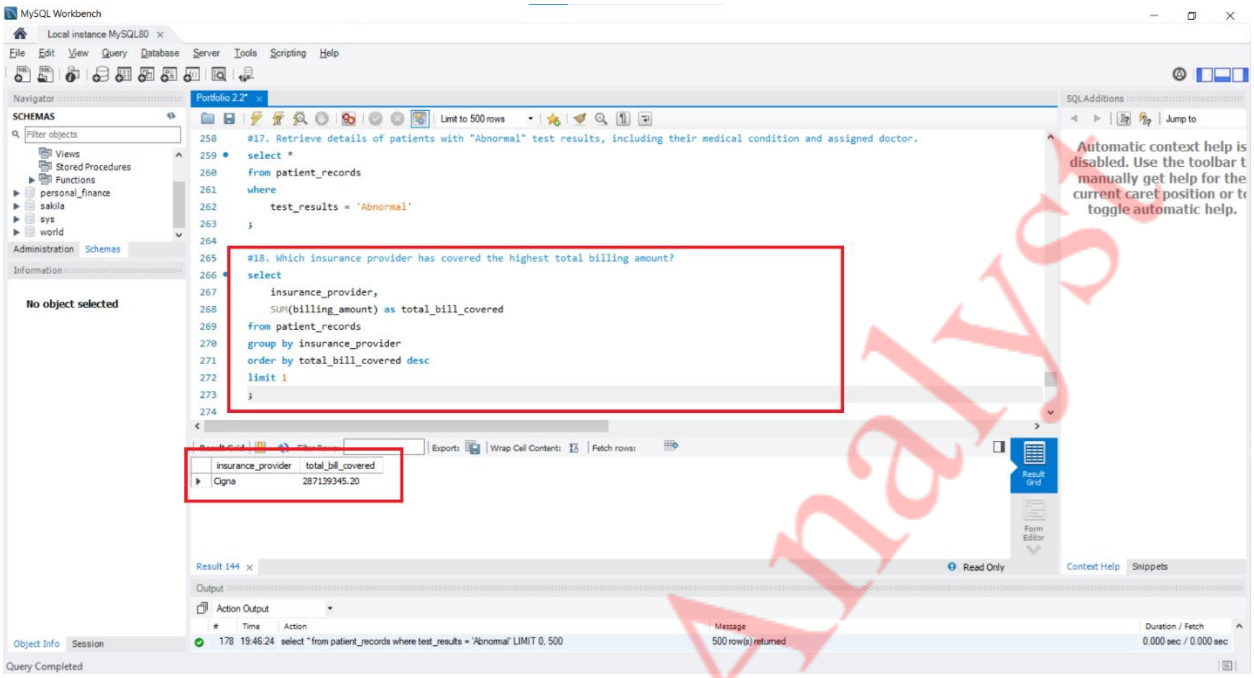
16. Identify the **most prescribed medication**, along with the number of times it was prescribed.



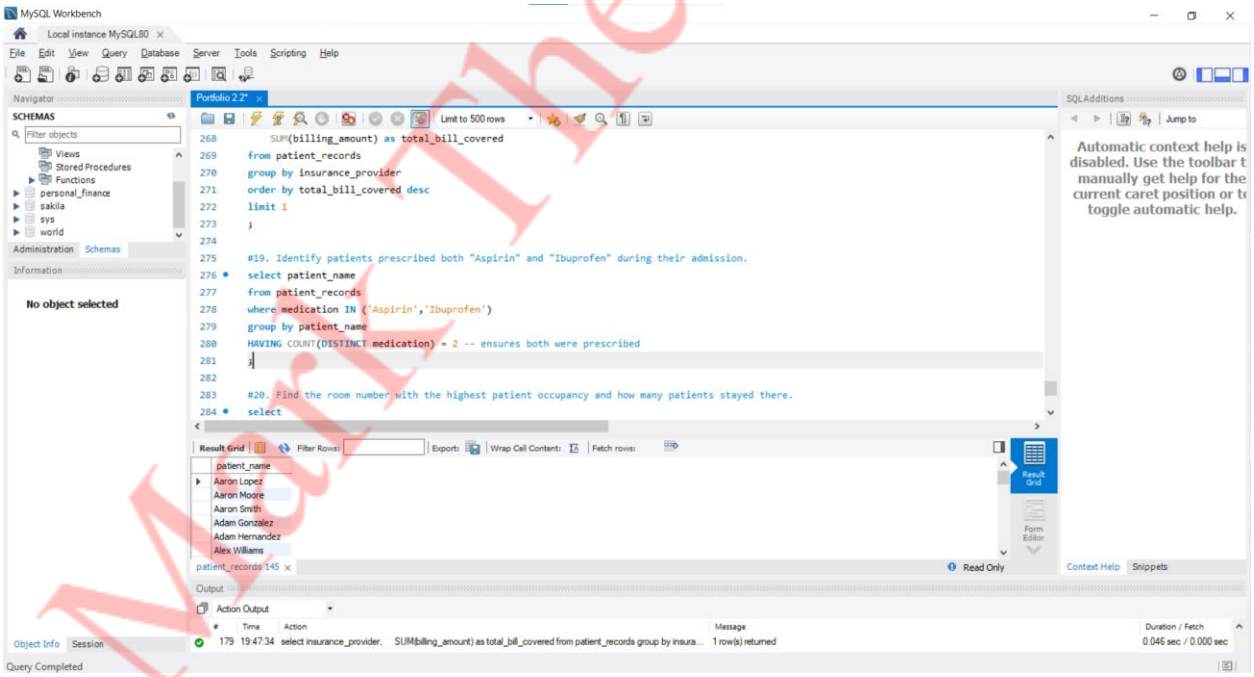
17. Retrieve details of **patients with "Abnormal" test results**, including their **medical condition and assigned doctor**.



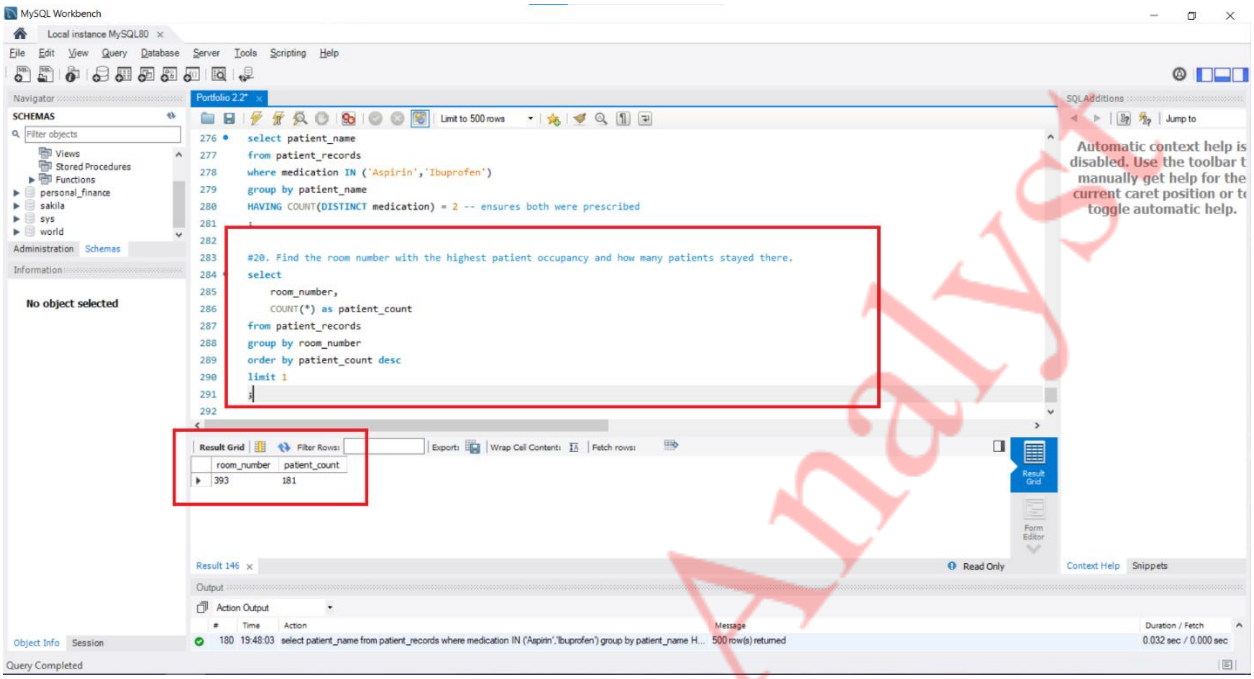
18. Which **insurance provider** has covered the **highest total billing amount**?



19. Identify **patients prescribed both "Aspirin" and "Ibuprofen"** during their admission.



20. Find the **room number with the highest patient occupancy** and how many patients stayed there.



KEY INSIGHTS

- The dataset includes **55,000 unique patients**, with a near-even gender split — **27,726 female** patients.
- **Admission types** are evenly distributed: **Urgent (18,576), Emergency (18,269), Elective (18,655)**.
- **LLC Smith Hospital** recorded the highest number of patient admissions (44), while **Johnson PLC** had the **highest total billing** at **₱1,084,202.70**.
- Among **medical conditions**, the most common are **Arthritis (9,308)**, **Diabetes (9,304)**, and **Hypertension (9,245)**.
- **Michael Smith** stands out as the doctor with the most treated patients (27).
- Patients stay in the hospital for an **average of 15.51 days**, and **978** admissions occurred in the last **30 days**.
- The **busiest month** was **August 2020** with **1,014 admissions**.
- **5,193 emergency admissions** took place during weekends, suggesting high weekend demand.
- The **most prescribed medication** appears over **11,140 times**, and **Cigna** was the top insurance provider by billing at **₱287.1M**.
- **Room 393** had the highest occupancy, accommodating **181 patients**.

RECOMMENDATIONS

- **Staffing & Resource Allocation:** Given the even distribution across admission types and frequent weekend emergencies, hospitals should ensure adequate weekend staffing and resources, especially for emergency cases.
- **Doctor Workload Monitoring:** With Michael Smith treating the most patients, reviewing doctor-patient ratios may help avoid potential burnout or improve efficiency.
- **Facility Usage:** High usage of Room 393 may point to availability issues or operational preference — deeper analysis could ensure better space optimization.
- **Medical Supply Planning:** Frequent prescriptions and top medical conditions (e.g., arthritis, diabetes) highlight areas to prioritize in medication inventory and chronic care programs.

- **Insurance & Billing Optimization:** Cigna's large billing coverage may indicate a strong partnership — but reviewing claim approval rates and processing times could help manage financial flows more effectively.

🔄 REFLECTION & NEXT STEPS

- This project sharpened my skills in writing complex SQL queries and analyzing real-world healthcare data. It gave me practical experience uncovering trends in patient care and billing. Moving forward, I plan to connect the data to a visualization tool to present insights more clearly and explore long-term patterns.