Upon successfully executing data integration operations in Python, I have generated five distinct datasets by combining Transactional and Reference data. These datasets are structured as follows:

Q1 Dataset: In this dataset, I have merged data from the accounts and diagnosis dataset with relevant information extracted from the look_up_diagnosis dataset. The purpose of this integration is to combine account and diagnosis data with diagnosis-specific details.

Q2 Dataset: This dataset is a combination of data from the Accounts dataset, Charges Dataset, and relevant details retrieved from the lookup_table_MS DRG. It aims to merge account and charge data with insights related to MS DRG codes.

Q3 Dataset: This dataset merges data sourced from the Accounts dataset with supplementary information obtained from the lookup_table_provider. It serves to combine account-related data with provider-specific details.

Q4 Dataset: By merging data from the Accounts dataset with information sourced from the lookup_table_facility, this dataset provides a comprehensive view that incorporates facility-specific attributes into the account data.

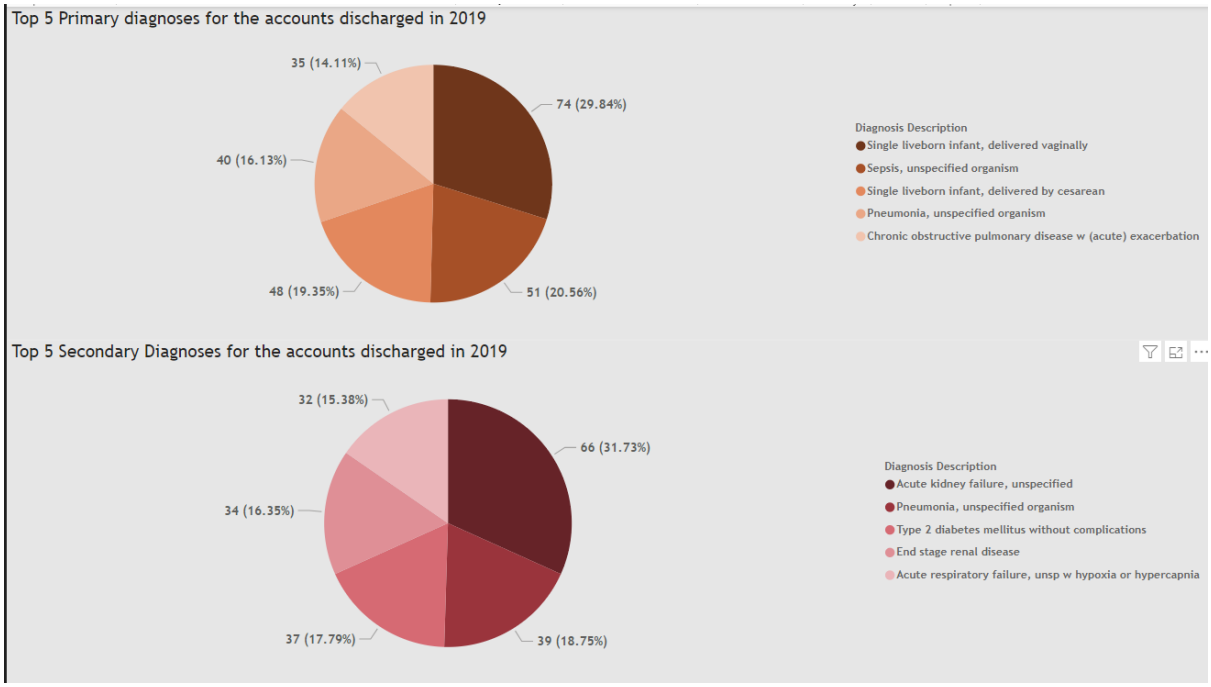
Q5 Dataset: This dataset represents a combination of data from the Accounts dataset, along with details extracted from both the lookup_table_MS DRG Service Line Description and the lookup_table_MS DRG. This integration provides a holistic perspective, incorporating MS DRG service line descriptions, account data, and MS DRG-related insights.

Answering the questions using MySQL and Power BI

1. For accounts discharged during 2019, which were the most common primary diagnoses? Secondary diagnoses?

```
SELECT *
FROM (
    SELECT
        `Diagnosis Description`,
        `Diagnosis Sequence`,
        `Dx Clinical Classification`,
        COUNT(*) AS `Frequency`,
        RANK() OVER (PARTITION BY `Diagnosis Sequence` ORDER BY COUNT(*) DESC) AS `Rank`
    FROM q1
    WHERE `Discharge Year` = '2019' AND (`Diagnosis Sequence` = 1 OR `Diagnosis Sequence` = 2)
    GROUP BY `Diagnosis Sequence`, `Diagnosis Description`, `Dx Clinical Classification`
) AS RankedData
WHERE `Rank` <=5
ORDER BY `Rank`, `Diagnosis Sequence`, `Frequency` DESC;
```

Diagnosis Description	Diagnosis Sequence	Dx Clinical Classification	Frequency	Rank
Single liveborn infant, delivered vaginally	1	Liveborn (218)	74	1
Acute kidney failure, unspecified	2	Acute and unspecified renal failure (157)	66	1
Sepsis, unspecified organism	1	Septicemia (except in labor) (2)	51	2
Pneumonia, unspecified organism	2	Pneumonia (except that caused by tuberculosis ...	39	2
Single liveborn infant, delivered by cesarean	1	Liveborn (218)	48	3
Type 2 diabetes mellitus without complications	2	Diabetes mellitus without complication (49)	37	3
Pneumonia, unspecified organism	1	Pneumonia (except that caused by tuberculosis ...	40	4
End stage renal disease	2	Chronic kidney disease (158)	34	4
Chronic obstructive pulmonary disease w (acute...	1	Chronic obstructive pulmonary disease and bro...	35	5
Acute respiratory failure, unsp w hypoxia or hy...	2	Respiratory failure; insufficiency; arrest (adult) ...	32	5

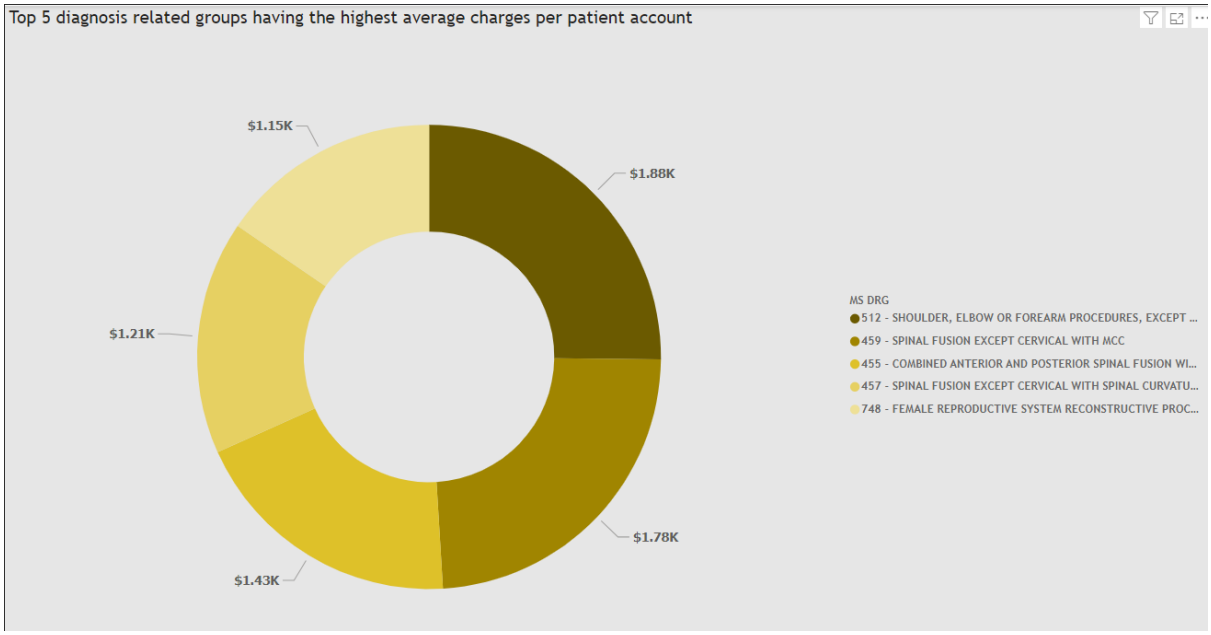


2. Which diagnosis-related groups (MS-DRGs) have the highest average charges per patient account?

```
SELECT `MS DRG`, a_id, AVG(total_charges) AS avg_total_charges
FROM q2
GROUP BY `MS DRG`, a_id
ORDER BY avg_total_charges DESC
LIMIT 10;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

MS DRG	MS DRG Description	avg_charges_per_account
512 - SHOULDER, ELBOW OR FOREARM PROCE...	SHOULDER, ELBOW OR FOREARM PROCEDURE...	1875.000000
459 - SPINAL FUSION EXCEPT CERVICAL WITH ...	SPINAL FUSION EXCEPT CERVICAL WITH MCC	1778.707286
455 - COMBINED ANTERIOR AND POSTERIOR S...	COMBINED ANTERIOR AND POSTERIOR SPINA...	1431.932066
457 - SPINAL FUSION EXCEPT CERVICAL WITH ...	SPINAL FUSION EXCEPT CERVICAL WITH SPIN...	1212.540705
748 - FEMALE REPRODUCTIVE SYSTEM RECON...	FEMALE REPRODUCTIVE SYSTEM RECONSTRUC...	1151.160526

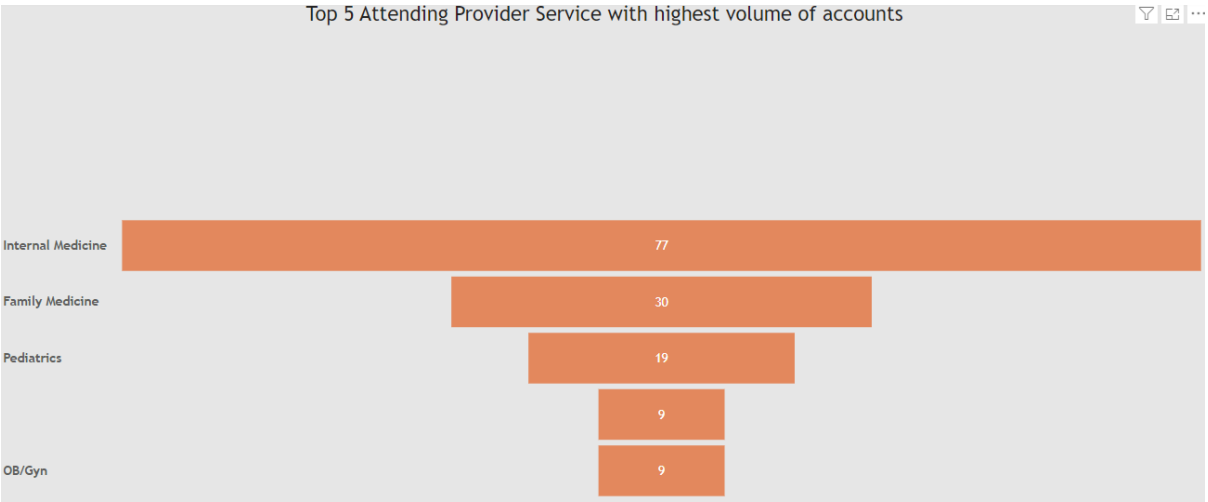


3. Which Attending Provider Service had the highest volume of accounts (encounters) during December 2018?

```
SELECT `Provider Service`, COUNT(DISTINCT a_id) AS encounter_count
FROM accounts_provider
WHERE YEAR(admit_date) = 2018 AND MONTH(admit_date) = 12
GROUP BY `Provider Service`
ORDER BY encounter_count DESC
LIMIT 5;
```

Result Grid | Filter Rows:

Provider Service	encounter_count
Internal Medicine	77
Family Medicine	30
Pediatrics	19
	9
OB/Gyn	9

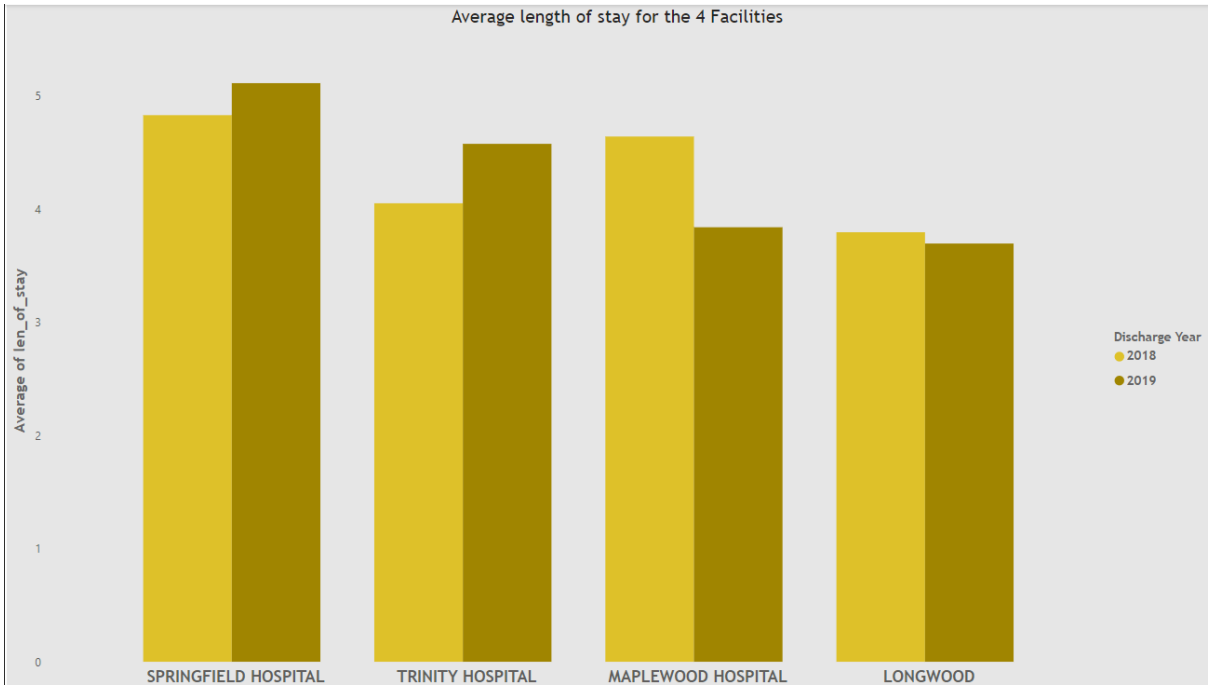


4. What is the "average length of stay" (number of hospital days between admit date and discharge date, but not including discharge date) for each of the 4 facilities for each of the years for which account information is provided (2018 & 2019).

```
select avg(DATEDIFF(`Discharge Date`, `Admit Date`)) AS average_length_of_stay, `Facility Name`,
`Discharge Year`
from q4
group by `Facility Name`, `Discharge Year`;
```

Result Grid | Filter Rows: | Export: | Wra

average_length_of_stay	Facility Name	Discharge Year
4.0519	TRINITY HOSPITAL	2018
4.5771	TRINITY HOSPITAL	2019
3.7958	LONGWOOD	2018
3.6961	LONGWOOD	2019
4.8305	SPRINGFIELD HOSPITAL	2018
5.1134	SPRINGFIELD HOSPITAL	2019
4.6417	MAPLEWOOD HOSPITAL	2018
3.8399	MAPLEWOOD HOSPITAL	2019



5. Which MS-DRG service lines had the largest difference between the observed length of stay and the national average provided in the MS-DRG reference file? Use the "Arithmetic Mean LOS" field as the national average.

```
select `MS DRG`, `Service Line Description`, `Observed LOS` - `Arithmetic Mean LOS` as difference
from q5
order by  difference desc
limit 5;
```

	MS DRG	Service Line Description	difference
▶	004 - TRACHEOSTOMY WITH MV >96 HOURS OR PRINCIPAL DIAGNOSIS EXCEPT FACE, MOUTH AND...	SURGICAL TRACHEOSTOM	76.4
	004 - TRACHEOSTOMY WITH MV >96 HOURS OR PRINCIPAL DIAGNOSIS EXCEPT FACE, MOUTH AND...	SURGICAL TRACHEOSTOM	49.4
	025 - CRANIOTOMY AND ENDOVASCULAR INTRACRANIAL PROCEDURES WITH MCC	NEUROSURGERY	38.2
	563 - FRACTURE, SPRAIN, STRAIN AND DISLOCATION EXCEPT FEMUR, HIP, PELVIS AND THIGH WIT...	OTHER ORTHOPAEDICS	34.5
	793 - FULL TERM NEONATE WITH MAJOR PROBLEMS	NEONATOLOGY	33.3

Top 5 Service Lines having the highest difference between the observed length of stay and the national average

MS DRG	Service Line Description	Difference
793 - FULL TERM NEONATE WITH MAJOR PROBLEMS	NEONATOLOGY	33.30
563 - FRACTURE, SPRAIN, STRAIN AND DISLOCATION EXCEPT FEMUR, HIP, PELVIS AND THIGH WITHOUT MCC	OTHER ORTHOPAEDICS	34.50
025 - CRANIOTOMY AND ENDOVASCULAR INTRACRANIAL PROCEDURES WITH MCC	NEUROSURGERY	38.20
004 - TRACHEOSTOMY WITH MV >96 HOURS OR PRINCIPAL DIAGNOSIS EXCEPT FACE, MOUTH AND NECK WITHOUT MAJOR O.R. PROCEDURES	SURGICAL TRACHEOSTOM	49.40
004 - TRACHEOSTOMY WITH MV >96 HOURS OR PRINCIPAL DIAGNOSIS EXCEPT FACE, MOUTH AND NECK WITHOUT MAJOR O.R. PROCEDURES	SURGICAL TRACHEOSTOM	76.40