**Group Project Report**

**Anita Adhikari, Brandon Kivlin, and JoMario Rivera**

**10/07/2023**

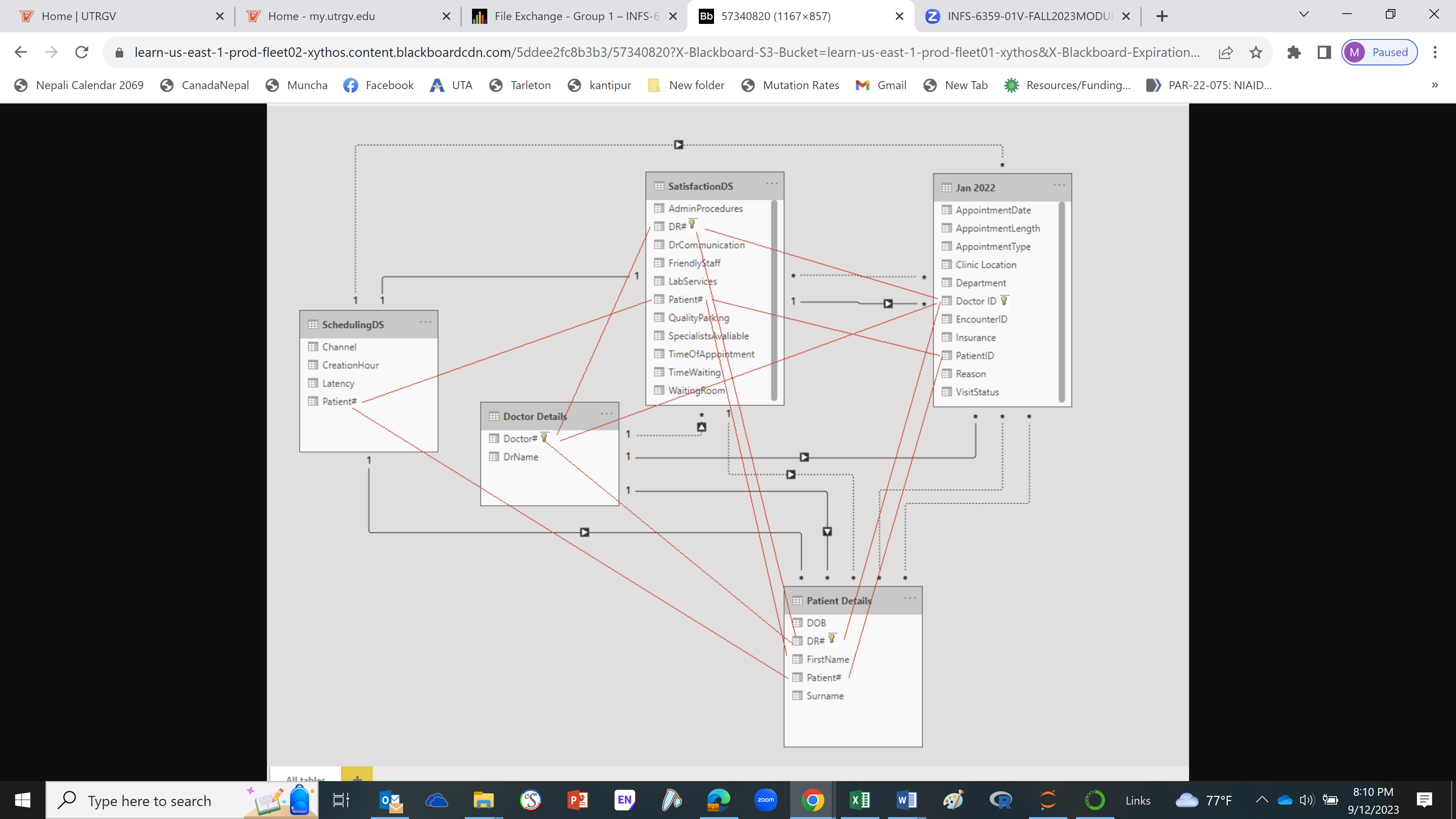
1. **Business background and decisions**: The Texas Valley Plus Clinic is a small medical clinic that offers various healthcare services in the Dallas, TX. The clinic has been committed to providing accessible, prompt, and professional healthcare service with a high professional standard of care for the patients. We aid with scheduling patient appointments, medical and mental health assessments,**and provide routine immunizations and referral services to specialty care when warranted. Similarly, we offer**prescription orders, billing and insurance claims processing, and patient feedback. **We offer basic advice regarding patients’ medical conditions and their health management. This facility**provides strong patient education and decision aids by understanding their cultural and personal preferences and engaging family and caregivers in their journey to healthy living.

 The Texas Valley Plus Clinic has data on patient appointments, medical treatments administered, insurance coverage, billing records, and patient feedback surveys. The major business decision that the system can support are listed below:

* The clinic desires insights into patient appointment scheduling patterns, common medical treatments administered, and patient demographics. This in turn helps generate insights and show trends and patterns **to proactively plan for potential resource allocation adjustments in response to the health trends found.**
* Analyzing the insurance coverage and billing data helps get insights in cost management.
* Analyzing the patient satisfaction scores data aids in evaluating business performance, discover new business opportunities, and make better business decisions related to our patients and services.

1. **Transactional database**: The organization has a dataset from the year 2022. It includes information on patient ID, appointment date, appointment type, length of appointment, reason for visit, clinic location, provider’s information, type of insurance, etc.  We analyzed the data using Excel and MySQL.

We have built a database comprised of patient details which includes information on patient’s first name, last name, and date of birth, patient ID number, and doctor’s name. Similarly, information about appointment creation (date, time, and method of appointment such as phone call, personal, or through web access), provider’s information, and patient satisfaction survey are collected and analyzed. The entity relationship between different datasets are shown below.



**Entity relationship diagram**

**PK-**DR#<->Doctor#<->DoctorID

**FK-**PatientID<->Patient#

1. **Data mart:**

Part 1: Star schema

**A screenshot of a computer

Description automatically generated**

Part 2: SQL Query

SELECT

PK\_PatientID,

AppointmentDate,

AppointmentType,

AppointmentLength,

Reason,

Insurance,

VisitStatus,

FirstName,

Surname,

DOB,

DrCommunication,

FriendlyStaff,

LabServices,

WaitingRoom,

QualityParking,

DrName,

CreationHour,

Latency,

Channel

FROM

AppointmentDS, DoctorDS, PatientDS, SatisfactionDS,SchedulingDS

WHERE

SatisfactionDS.FK\_PatientID = PatientDS.PK\_PatientID

AND SatisfactionDS.FK\_AppointmentID = AppointmentDS.PK\_AppointmentID

AND SatisfactionDS.FK\_DoctorID = DoctorDS.PK\_DoctorID

AND SchedulingDS.FK\_PatientID = PatientDS.PK\_PatientID

AND SchedulingDS.FK\_AppointmentID = AppointmentDS.PK\_AppointmentID;

1. **OLAP cube**

We used the online analytical processing (OLAP) tool to quickly and easily retrieve the information from our data for analysis purpose.

Measure: We analyzed the sum of patient appointment based on different insurance types (Medicare, Private, or Cash pay).

Dimension: The analysis was based on appointments scheduled by quarter.

Granularity: in order to obtain a higher level of granularity we created a date hierarchy based on Date, Quarter, and Month attributes from the dataset.

A screenshot of a computer

Description automatically generated

Based on the hierarchy, we developed some pivot tables. A screenshot of a graph

Description automatically generated

From the above pivot table, we can see that across the year, a total of 1,645 appointments were scheduled. Private payments are the dominant method with 1,108 appointments. Medicare covers 504 appointments, indicating it's the primary government health coverage used. Cash pay sees the lowest usage with only 33 appointments. Some trends seen is that there is an increasing trend in Private payments from Q1 through Q3 suggests a growing customer base or increased offerings that appeal to those with private insurance. A possible strategy to implement is to understand the reason for the dominance of private payments, which can help refine marketing strategies and partnerships with insurance companies.

A screenshot of a computer screen

Description automatically generated

The Pivot table shows the count of appointments based on the types of patients. For follow-up Visit, the highest number of patients scheduled their appointments during quarter 4. Whereas, during quarter 3, the highest number of patients scheduled their appointment as a New Patient. For Procedure, the highest number of patients scheduled their appointments during the first quarter. From this we can say that there is an increase in number of appointments for follow-up visit towards the end of the year which is exactly the opposite for Procedure visits.

In terms of monthly patterns, January, April, and July started their respective quarters with higher counts, suggesting a possible trend of more appointments being booked at the beginning of each quarter. There’s consistent attendance for "Follow-up Visits" each month, ranging from 70 (lowest in February) to 108 (highest in May). "New Patient" visits peaked in July with 50 appointments and reached their lowest in January with 25. "Procedures" were relatively infrequent, with the highest being 11 in March and several months having as few as 3 or 4.

In terms of appointment type patterns, "Follow-up Visits" consistently outnumber other types of appointments every month. "New Patient" visits fluctuate monthly but maintain a steady presence. "Procedures" are the least common type of appointment, never surpassing 20 in any month.

Finally, some potential implications are that there is a consistent number of "Follow-up Visits" indicates that there's a steady stream of returning patients. The variable number of "New Patient" appointments suggests there might be external factors affecting new patient registrations or it could be driven by outreach or advertising efforts.

A screenshot of a data table

Description automatically generated

Above Pivot table shows the distribution of appointment based on clinic location. Brentwood clinic has highest number of appointments in the months of January, March, April, July, September, and October. Similarly, Culver City clinic has highest number of appointments during January, July, and December and Marina Del Rey clinic in February, May, and August. Moreover, Santa Monica clinic has highest number of appointments in June and November and Venice location in May and July. From this we can say that Brentwood location remains comparatively busy throughout the year with a greater number of patient appointments as compared to other locations. This data helps us to allocate resources and staff to those locations with high patient volumes which in turn helps with reduced waiting times, quality patient care, and increased patient satisfaction.

1. **Mining structure:** We have a dataset including 1645 records on patients which includes categorical variables such as Insurance (Medicare, Private, or Cash), Appointment Type (New patient, Follow-up, or Procedure),and Visit Status (Completed or No Shows). In this study, we are trying to predict Visit status with relation to input variables Insurance and Appointment type**.**

We performed data mining using the decision tree method. The predicted variable is appointment status. This is the outcome we are trying to predict based on the decision tree. We want to see whether an appointment was completed or not. This is a discrete variable, as it has distinct categories (Completed or Not Completed).

The input variables are insurance type and the nature of the appointment. From the tree, we can see that there are three distinct branches/categories for insurance type: Medicare, Private, or Cash Pay. This is a discrete variable as it represents specific categories. Nature of appointment refers to whether the appointment was for a procedure, a follow-up visit, or if it was a new patient's visit. Again, this is a discrete variable because it is a distinct category.

**A diagram of a company

Description automatically generated with medium confidence**

From the above decision tree, we can see that patients with private insurance or Medicaid coverage have higher instances of showing up for their appointments. Whereas patients who pay cash and new patients have 87.5% instances that they do not show up for the appointment. Similarly, patients who pay cash and Follow-up appointments have 47.6% instances of not showing up. We can suggest offering affordable payment options such as payment in installments for those patients who do not have insurance coverage and want to pay cash for their doctor’s visit. This will reduce the no-shows as well as ensure that the patients don’t have to struggle to pay their medical bills.