**FINAL REPORT**  
**Medical Part D Dataset**

Project is about Medicare Part D which is an optional United States federal-government program to help Medicare beneficiars pay for self-administered prescription drugs.  
  
**Changes/Additions**   
Updated the tables in the ER diagram so it suits rdf structure, normalized and divided data. We also added IDs to drug table.  **Data Loading Process**   
 After doing normalization, divided the raw data into different csv files so have every column in the data separately (except drug part, take 3 columns associated with the drug and hold these 3 columns into another csv file). First, insert values into drug table directly from 3 column drug csv file. After that, placed the remaining column names as insertions in columns table. In the value table, used code to determine the column IDs and years for each CSV file, associating them with the corresponsing drug IDs. After all, we managed to load all the data. The code example use is below:  
  
  
SET @rownum := 0;

LOAD DATA INFILE 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\Tot\_Spndng\_2017.csv'

INTO TABLE valuetable

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

(@value)

SET columns\_id = 1,

year = 2017,

Drug\_drugId = @rownum := @rownum + 1,

Value = CASE

WHEN @value = '' OR @value IS NULL THEN NULL

WHEN @value = '""' THEN NULL

ELSE CAST(REPLACE(@value, '""', '') AS DOUBLE)

END;  
**The link of the raw data:**

[**https://catalog.data.gov/dataset/medicare-part-d-spending-by-drug-401d2**](https://catalog.data.gov/dataset/medicare-part-d-spending-by-drug-401d2)

**View and Stored Procedures  
  
AvgOutlierFlagByYear (View):** This view is for getting average outlier flag amount for each year.  
  
  
  
  
**ManWithOutlierOne (View):** This is for getting manifacturers which have value 1 on outlier flag section.  
Output:  
  
metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu  
  
**SandozDrugs (View):** This is for getting all the drug information of Sandoz brand.  
  
 **GetAvgBenefByBrand (Stored Procedure)**: This stored procedure is getting a brand name as IN, and giving an average value as OUT. So, it is for getting average total beneficiars amount based on a specific brand name.  
Output:  
  
metin, yazı tipi, ekran görüntüsü, beyaz içeren bir resim

Açıklama otomatik olarak oluşturuldu

**GetMaxCAGR (Stored Procedure):** This is for getting the maximum value of the constant average change in spending per dosage unit. This stored procedure is with OUT, so it gives maximum value as an output.  
  
**ErtaczoOutlierbyYear(Stored Procedure):** This is for getting outlier flag value of Ertaczo brand for specified year in the parameter. It takes year as an IN parameter.  
  
  
  
  
**System Limitations** The raw data has too many integer values and it keeps the data obtained from many years as a new attribute were challenging for us. It could have been more string values and the data from multiple years could have been stored more efficiently. It would be a lot easier.

**Relational Table Specification**  
Have 3 tables named Drug, Columns and ValueTable.  
  
**Drug’s attributes are:**  
**-drugId (primary key)(INT):** Uniquely identifies drugs, increments by 1 each time.  
**-brandName (VARCHAR):** Brand name of the drug   
**-genericName (VARCHAR):** Ingredients of the drug  
**-manifacturerName (VARCHAR):** Manifacturer name of the drug  
  
**Columns’ attributes are:**  
**-id (primary key)(INT):** Uniquely identifies column names.  
**-columnName (VARCHAR):** Name of the column.  
  
**ValueTable’s attributes are:  
-Drug.drugId (foreign key) (primary key) (INT)  
-Columns\_id (foreign key) (primary key) (INT)  
-Year (primary key) (INT):** Every column repeats for different years, so there is a year attribute. **-value (DOUBLE):** We divided every column in data into different csv files, so this is the section where these data entries been kept.   
 **Codes**  
Pages of codes for loading data, so only included 2-3 pages. The other codes for loading the data have very similar syntax to below ones.

SET @rownum := 0;

LOAD DATA INFILE 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\Tot\_Spndng\_2017.csv'

INTO TABLE valuetable

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

(@value)

SET columns\_id = 1,

year = 2017,

Drug\_drugId = @rownum := @rownum + 1,

Value = CASE

WHEN @value = '' OR @value IS NULL THEN NULL

WHEN @value = '""' THEN NULL

ELSE CAST(REPLACE(@value, '""', '') AS DOUBLE)

END;

SET @rownum := 0;

LOAD DATA INFILE 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\Tot\_Spndng\_2018.csv'

INTO TABLE valuetable

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

(@value)

SET columns\_id = 1,

year = 2018,

Drug\_drugId = @rownum := @rownum + 1,

Value = CASE

SET @rownum := 0;

LOAD DATA INFILE 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\Tot\_Dsg\_Unts\_2018.csv'

INTO TABLE valuetable

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

(@value)

SET columns\_id = 2,

year = 2018,

Drug\_drugId = @rownum := @rownum + 1,

Value = CASE

WHEN @value = '' OR @value IS NULL THEN NULL

WHEN @value = '""' THEN NULL

ELSE CAST(REPLACE(@value, '""', '') AS DOUBLE)

END;

SET @rownum := 0;

LOAD DATA INFILE 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\Tot\_Dsg\_Unts\_2019.csv'

INTO TABLE valuetable

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

(@value)

SET columns\_id = 2,

year = 2019,

Drug\_drugId = @rownum := @rownum + 1,

Value = CASE

WHEN @value = '' OR @value IS NULL THEN NULL

WHEN @value = '""' THEN NULL

ELSE CAST(REPLACE(@value, '""', '') AS DOUBLE)

END;  
  
**Some Stored Procedure/View Codes**DELIMITER //

CREATE PROCEDURE GetAvgBenefByBrand(IN brandName VARCHAR(45), OUT avgValue DECIMAL(65,2))

BEGIN

SELECT avg(v.value) INTO avgValue

FROM drug d

JOIN valuetable v ON v.Drug\_drugId = d.drugId

WHERE d.brandName = brandName AND v.Columns\_id = 4;

END //

DELIMITER  
  
DELIMITER //

CREATE PROCEDURE ErtaczoOutlierbyYear(IN inputYear INT)

BEGIN

SELECT v.\*, d.brandname, d.genericname, d.manifacturername

FROM valuetable v

JOIN drug d ON v.Drug\_drugId = d.drugid

WHERE v.year = inputYear AND d.brandName = "Ertaczo" and Columns\_id = 8;

END //

DELIMITER ;