

# Practicum II CS5200

Zongyu Wu

Spring 2024

## Connect to database

```
# 1. Library
library("RMariaDB")
library("DBI")
library(ggplot2)

# 2. Settings
db_user <- 'admin'           # use your value from the setup
db_password <- 'wuzongyu'    # use your value from the setup
db_name <- 'practicum1'      # use your value from the setup

db_host <- 'cs5200.ceko8sdfannu.us-east-1.rds.amazonaws.com' # for aws
db_port <- 3306

# 3. Connect to remote server database

db.aws <- DBI::dbConnect(RMariaDB::MariaDB(), user = db_user, password = db_password,
                        dbname = db_name, host = db_host, port = db_port)

# 4. Knit setting
knitr::opts_chunk$set(connection = "db.aws", echo = FALSE, warning = FALSE, message = FALSE)
```

Analytical Query I: Top five sales reps with the most sales broken down by year. So, for each year for which there is data, the top five reps for that year. Think about how you would best present this information so it is easily usable by an analyst or manager.

##	rep_name	total_amount_sold	year
## 1	Walison da Silva	2277190	2020
## 2	Aneeta Kappoorthy	1517438	2020
## 3	Lynette McRowe	1354725	2020
## 4	Prasad Patel	1301593	2020
## 5	Helmut Schwab	749201	2020
## 6	Walison da Silva	4555540	2021
## 7	Lynette McRowe	3635264	2021
## 8	Aneeta Kappoorthy	3016620	2021
## 9	Prasad Patel	1810407	2021
## 10	Helmut Schwab	1291606	2021
## 11	Walison da Silva	2188879	2022
## 12	Aneeta Kappoorthy	1163873	2022
## 13	Prasad Patel	1020732	2022

```
## 14    Lynette McRowe          907841 2022
## 15     Veronika Sixt          779036 2022
```

Above shows the total amount sold for the top 5 sales.

**Analytical Query II: Total sold per product per quarter. Think about how to best display this information.**

Table 1: Displaying records 1 - 10

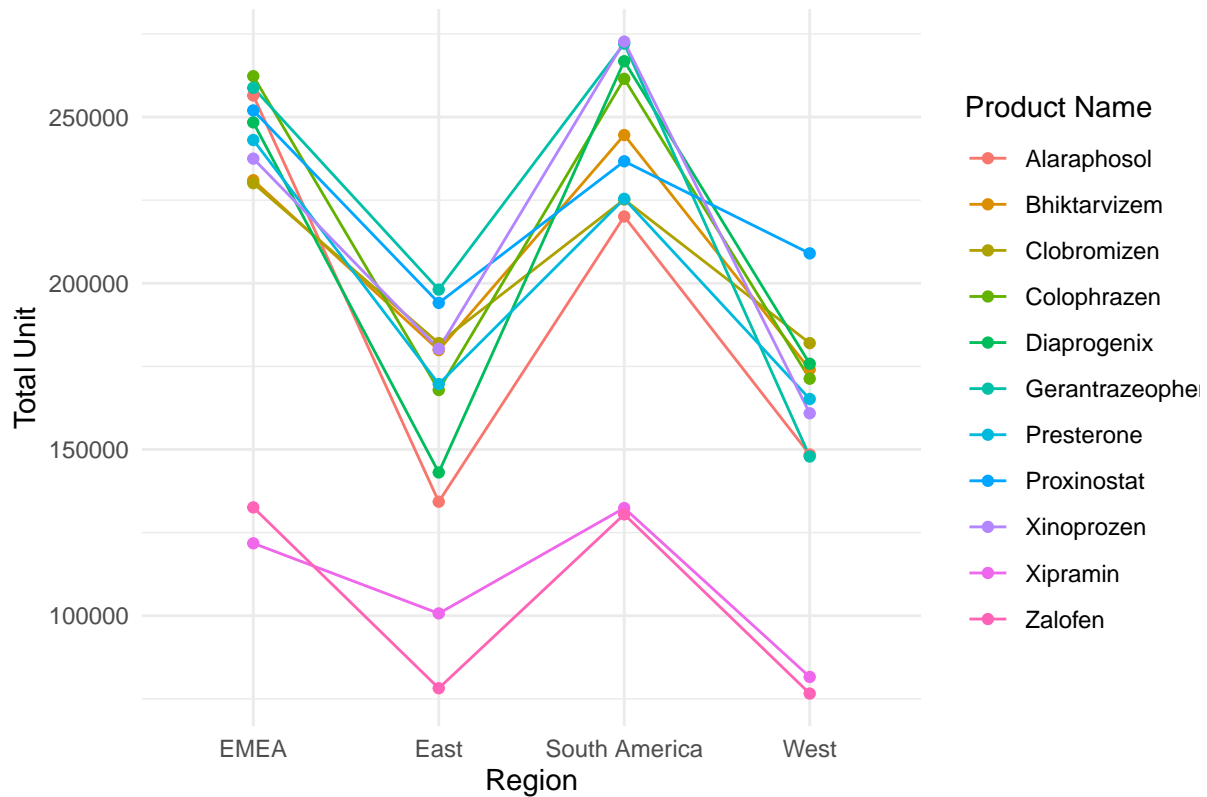
year	quarter	product_name	total_amount_sold
2020	1	Alaraphosol	53176
2020	2	Alaraphosol	45724
2020	3	Alaraphosol	56764
2020	4	Alaraphosol	34776
2020	1	Bhiktarvizem	182277
2020	2	Bhiktarvizem	254646
2020	3	Bhiktarvizem	271287
2020	4	Bhiktarvizem	168345
2020	1	Clobromizen	60048
2020	2	Clobromizen	104112

**Analytical Query III: Number of units sold per product per region. Show the result in a line graph visualization.**

```
##      product_name      region total_unit
## 1      Alaraphosol      EMEA    256500
## 2      Alaraphosol      East    134300
## 3      Alaraphosol South America    220100
## 4      Alaraphosol      West    148500
## 5      Bhiktarvizem      EMEA    231000
## 6      Bhiktarvizem      East    179900
## 7      Bhiktarvizem South America    244600
## 8      Bhiktarvizem      West    174100
## 9      Clobromizen      EMEA    230100
## 10     Clobromizen      East    182000
## 11     Clobromizen South America    225200
## 12     Clobromizen      West    182000
## 13     Colophrazen      EMEA    262300
## 14     Colophrazen      East    167900
## 15     Colophrazen South America    261500
## 16     Colophrazen      West    171300
## 17     Diaprogenix      EMEA    248400
## 18     Diaprogenix      East    143100
## 19     Diaprogenix South America    266800
## 20     Diaprogenix      West    175800
## 21 Gerantrazeophem      EMEA    258800
## 22 Gerantrazeophem      East    198100
## 23 Gerantrazeophem South America    272100
## 24 Gerantrazeophem      West    147900
## 25      Presterone      EMEA    243100
## 26      Presterone      East    169700
## 27      Presterone South America    225400
```

## 28	Presterone	West	165200
## 29	Proxinostat	EMEA	252000
## 30	Proxinostat	East	194100
## 31	Proxinostat	South America	236700
## 32	Proxinostat	West	209000
## 33	Xinoprozen	EMEA	237500
## 34	Xinoprozen	East	180300
## 35	Xinoprozen	South America	272700
## 36	Xinoprozen	West	160900
## 37	Xipramin	EMEA	121800
## 38	Xipramin	East	100700
## 39	Xipramin	South America	132400
## 40	Xipramin	West	81600
## 41	Zalofen	EMEA	132600
## 42	Zalofen	East	78200
## 43	Zalofen	South America	130500
## 44	Zalofen	West	76600

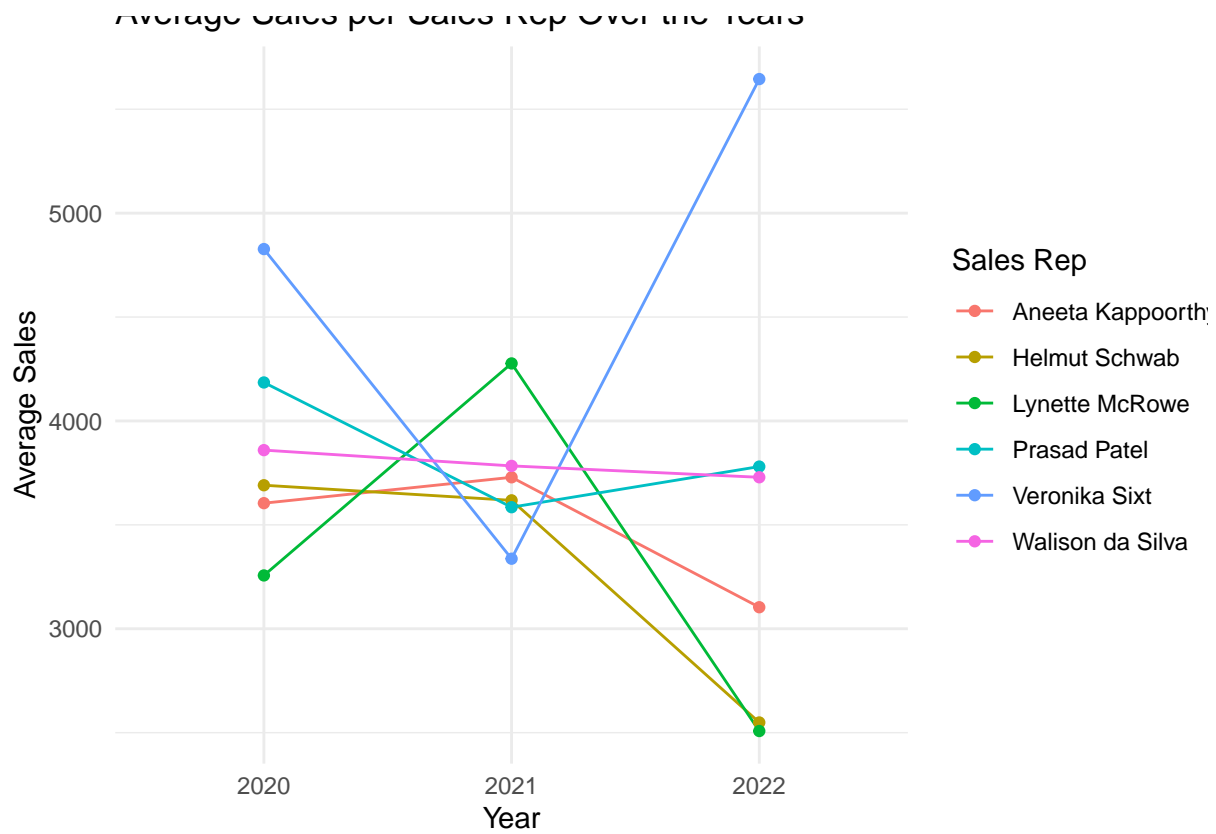
Total Unit by Region for Each Product



Analytical Query IV: Average sales per sales rep over the years. Show the result in a line graph visualization.

##	rep_name	avg_sales	year
## 1	Aneeta Kappoorthy	3604.366	2020
## 2	Aneeta Kappoorthy	3728.826	2021
## 3	Aneeta Kappoorthy	3103.661	2022
## 4	Helmut Schwab	3690.645	2020
## 5	Helmut Schwab	3617.944	2021

## 6	Helmut Schwab	2549.050	2022
## 7	Lynette McRowe	3256.550	2020
## 8	Lynette McRowe	4276.781	2021
## 9	Lynette McRowe	2507.848	2022
## 10	Prasad Patel	4185.186	2020
## 11	Prasad Patel	3584.964	2021
## 12	Prasad Patel	3780.489	2022
## 13	Veronika Sixt	4827.129	2020
## 14	Veronika Sixt	3337.123	2021
## 15	Veronika Sixt	5645.188	2022
## 16	Walison da Silva	3859.644	2020
## 17	Walison da Silva	3783.671	2021
## 18	Walison da Silva	3728.925	2022



Clean up