GROUP X

EDA PROJECT

INTRODUCTION

What is SQL?

Programming Language

Relational databases

Structured Language

SQL

Query

SQL is an acronym for Structured Query Language

SQL is the standard database language used by all RDBMS

MySQL

ORACLE

Adopted by

ANSI standard in 1986

ISO standard in 1987

Query language

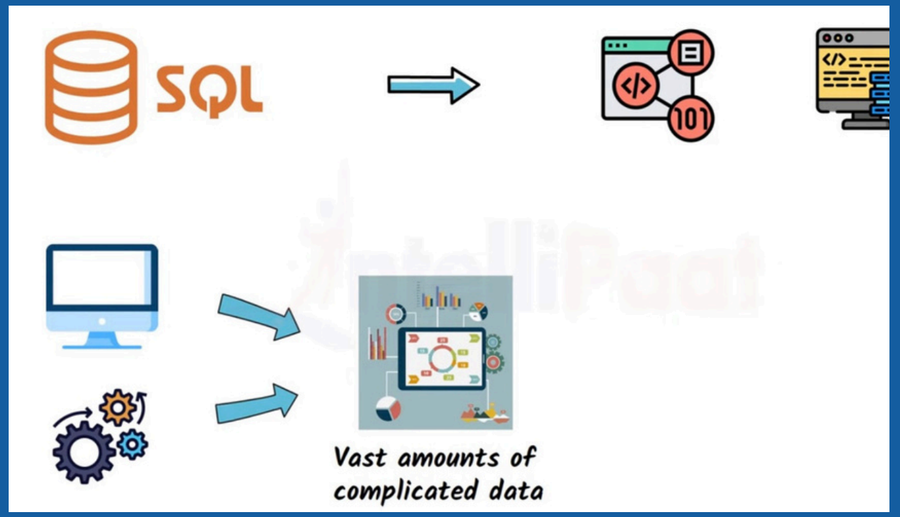
SQL is used by big companies

facebook

Instagram

LinkedIn

Used store data in the Back-end



Applications of SQL:

SELECT: used to retrieve data from a database

INSERT: used to insert data into a database

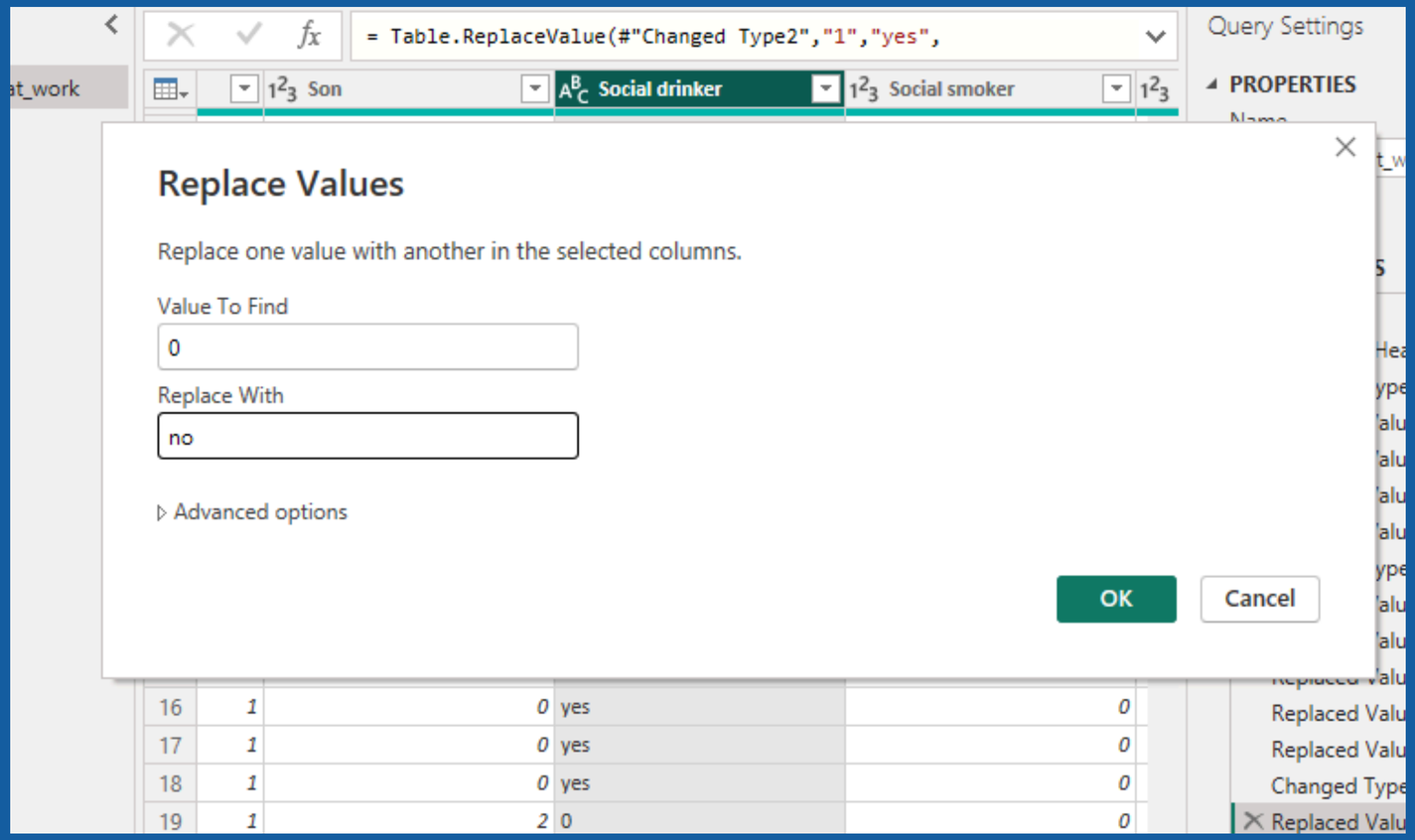
UPDATE: used to update existing data in a database

DELETE: used to delete data from a database

CREATE: used to create new tables, indexes, and other database objects

DROP: used to remove tables, indexes, and other database objects

Micrsoft Power BI

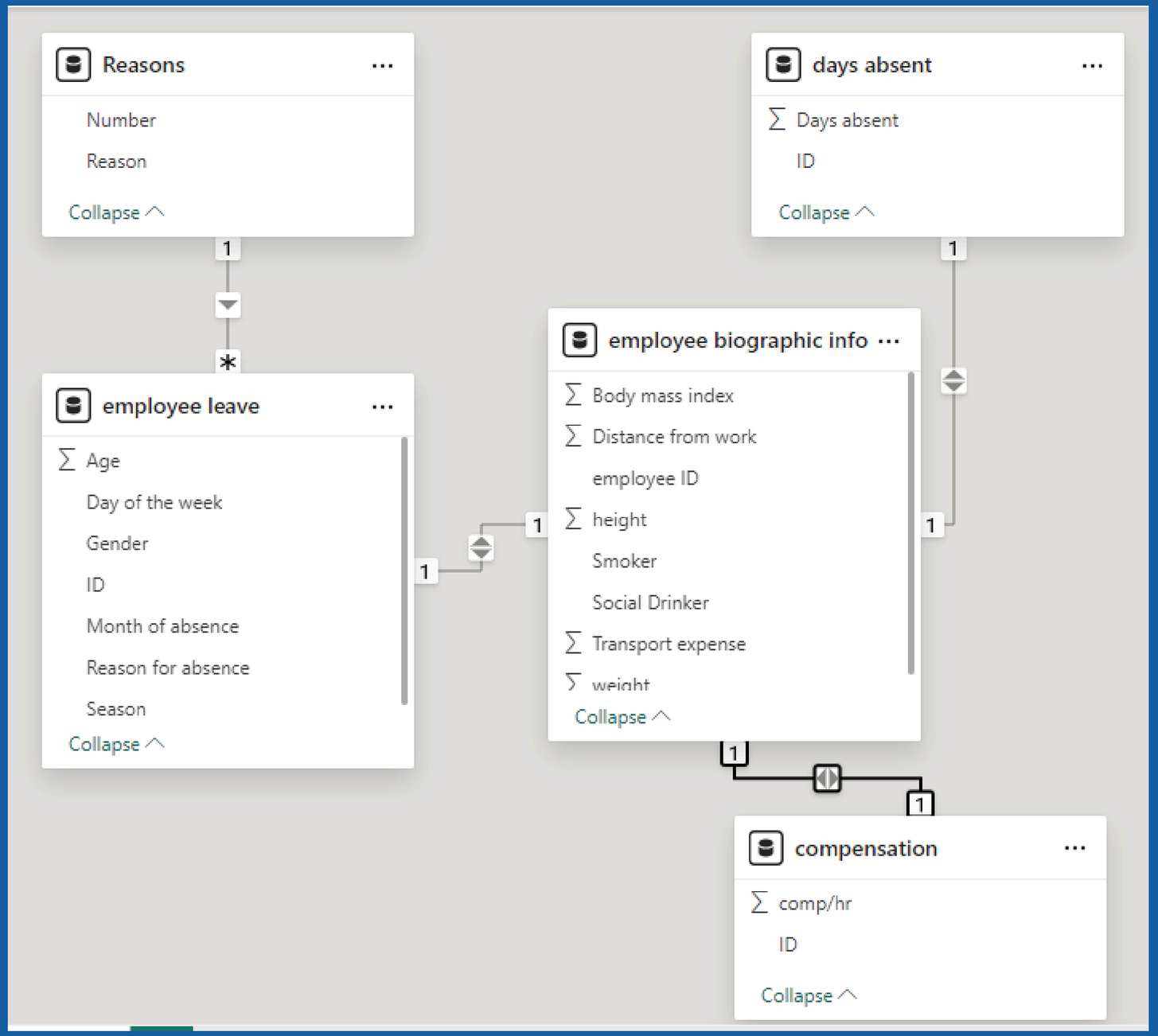


- compensation
  - comp/hr
  - ID
- days absent
  - Days absent
  - ID
- employee biographic info
  - Body mass index
  - Distance from work
  - employee ID
  - height
  - Smoker
  - Social Drinker
  - Transport expense
  - weight
- employee leave
  - Age
  - Day of the week
  - Gender
  - ID
  - Month of absence
  - Reason for absence
  - Season
- Reasons
  - Number
  - Reason

Data Analytics Tools

- Spreadsheet
- R
- Python
- Tableau

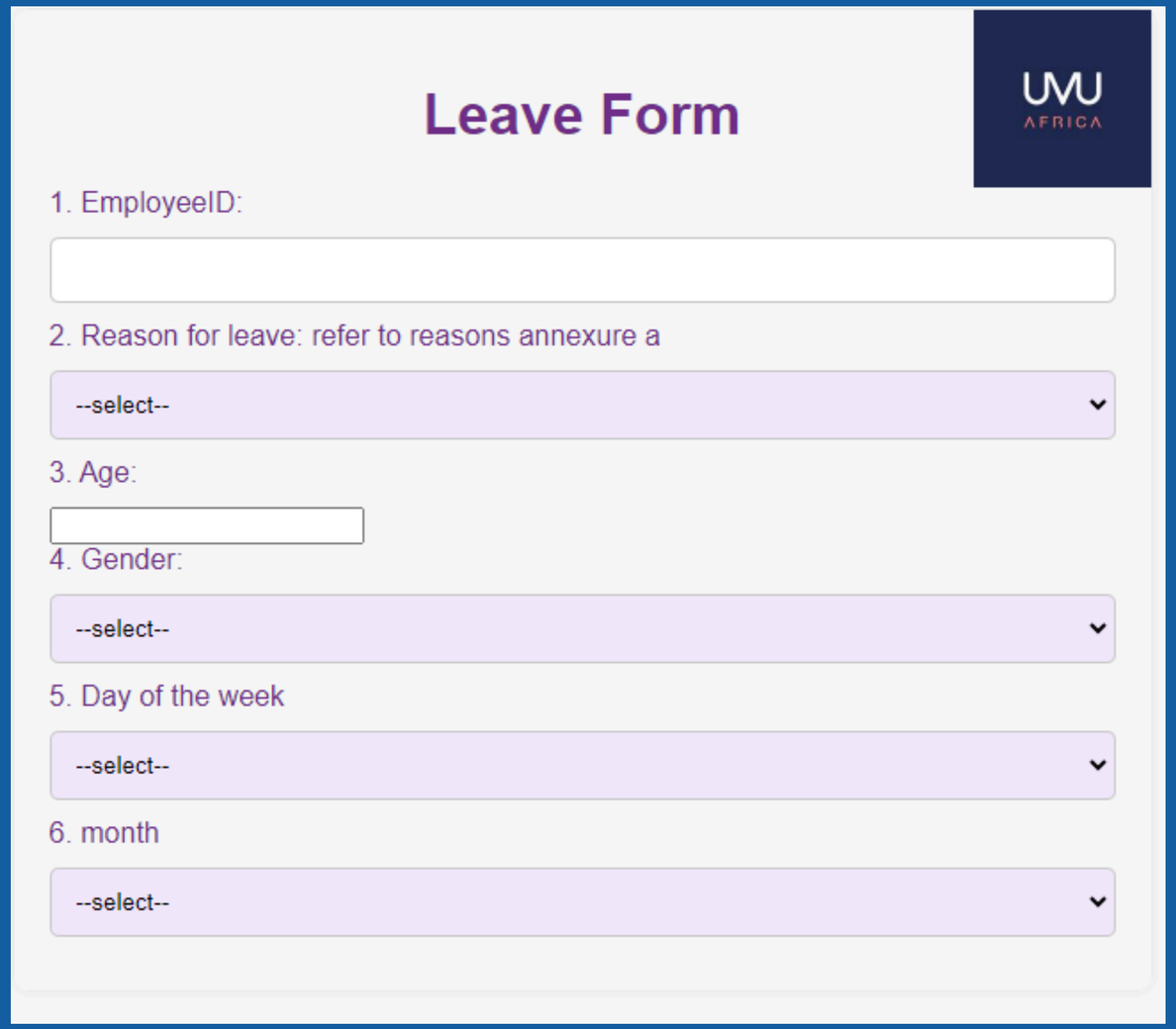
DATA STORAGE



- CSV
- DATABASES
- RELATIONAL
- NON RELATIONAL

PROJECT FILES AND DISCOVER

DATA COLLECTION



Leave Form

UVU AFRICA

1. EmployeeID:

2. Reason for leave: refer to reasons annexure a

--select--

3. Age:

4. Gender:

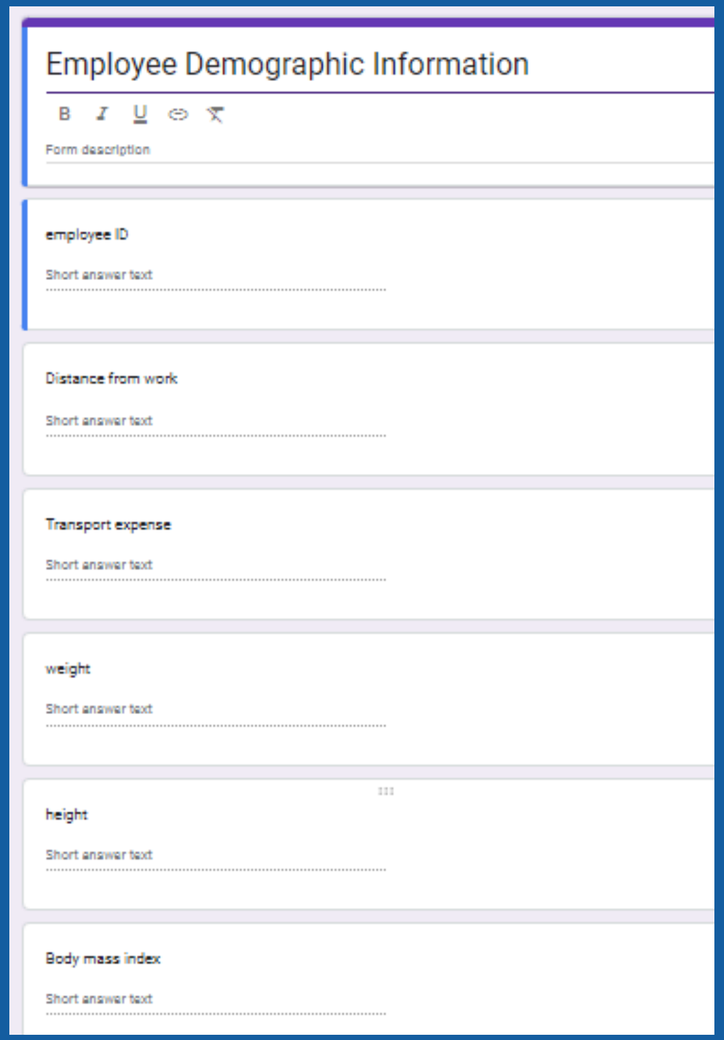
--select--

5. Day of the week

--select--

6. month

--select--



Employee Demographic Information

B I U

Form description

employee ID

Short answer text

Distance from work

Short answer text

Transport expense

Short answer text

weight

Short answer text

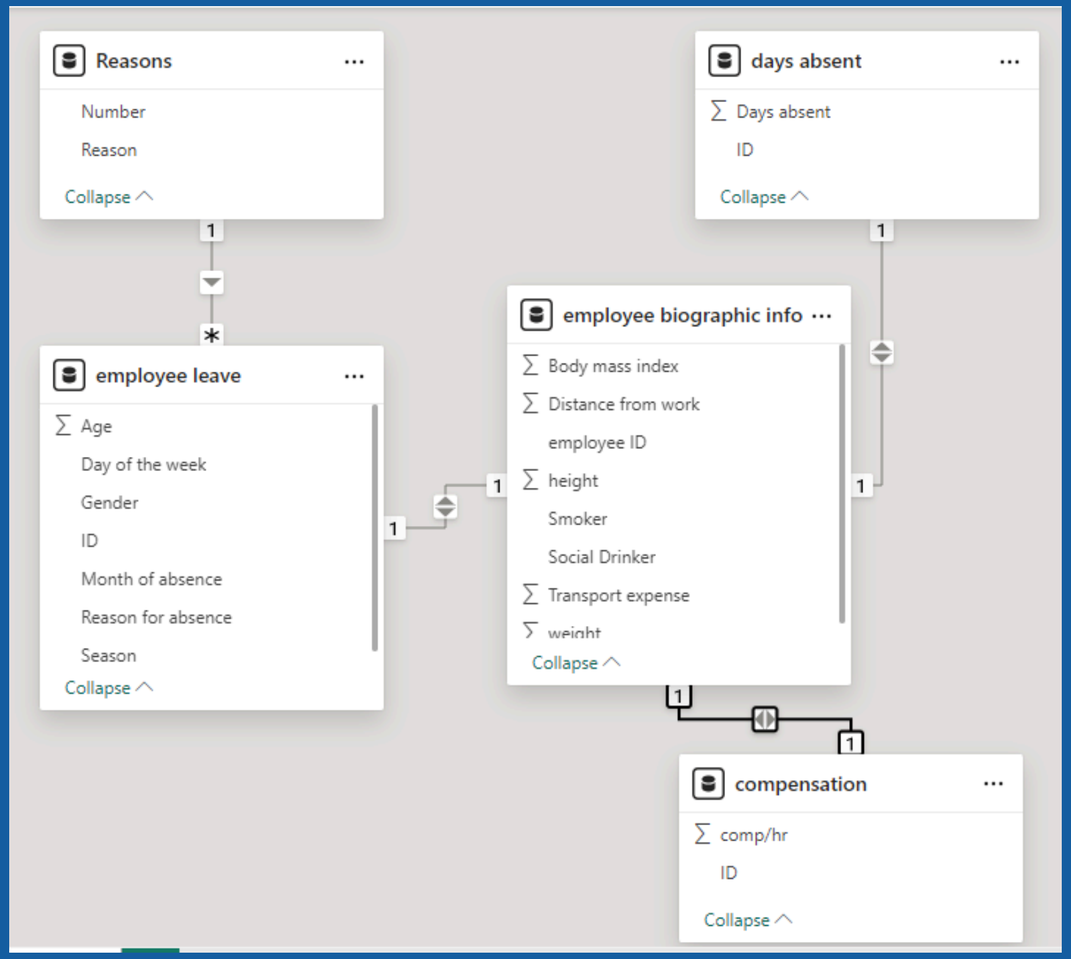
height

Short answer text

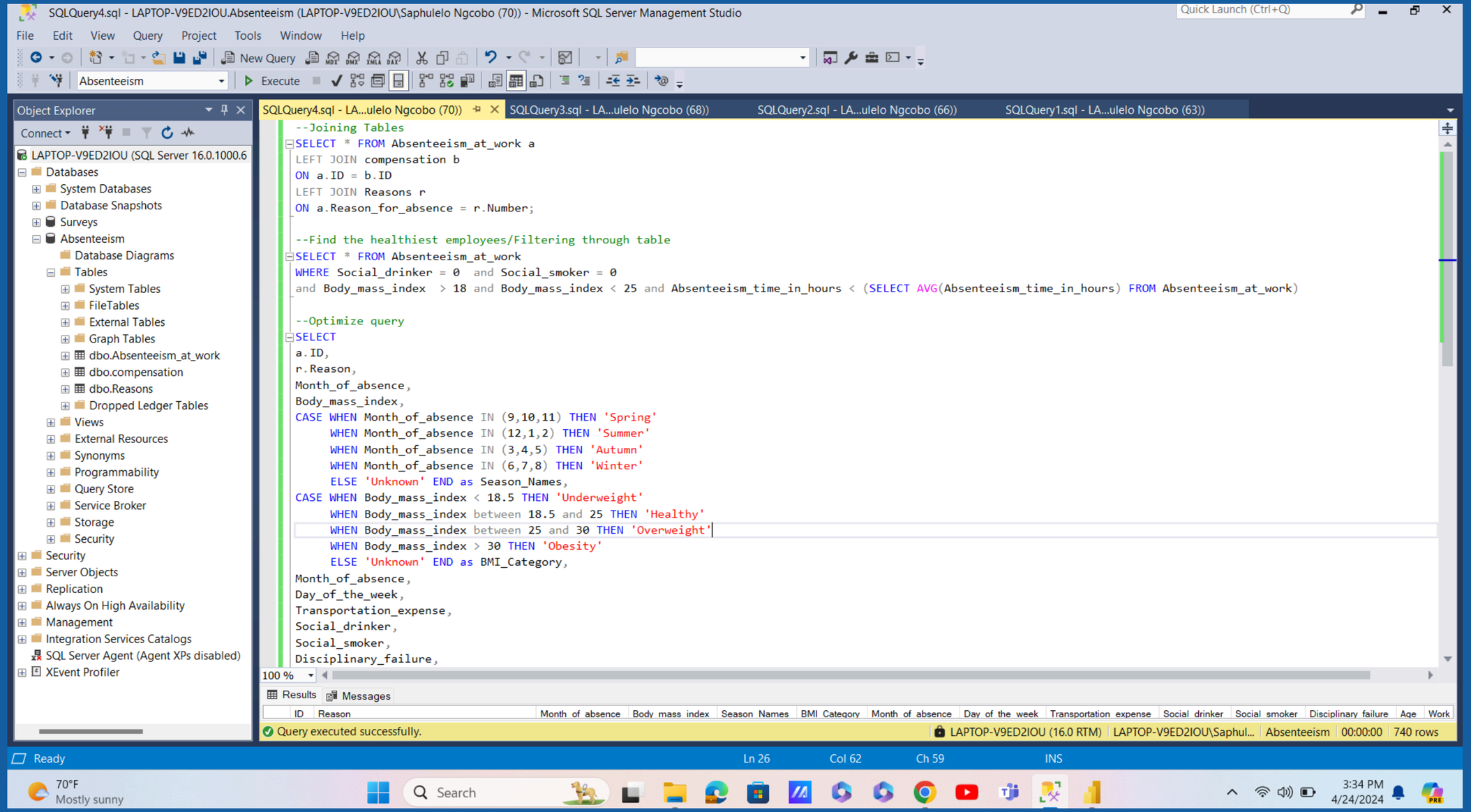
Body mass index

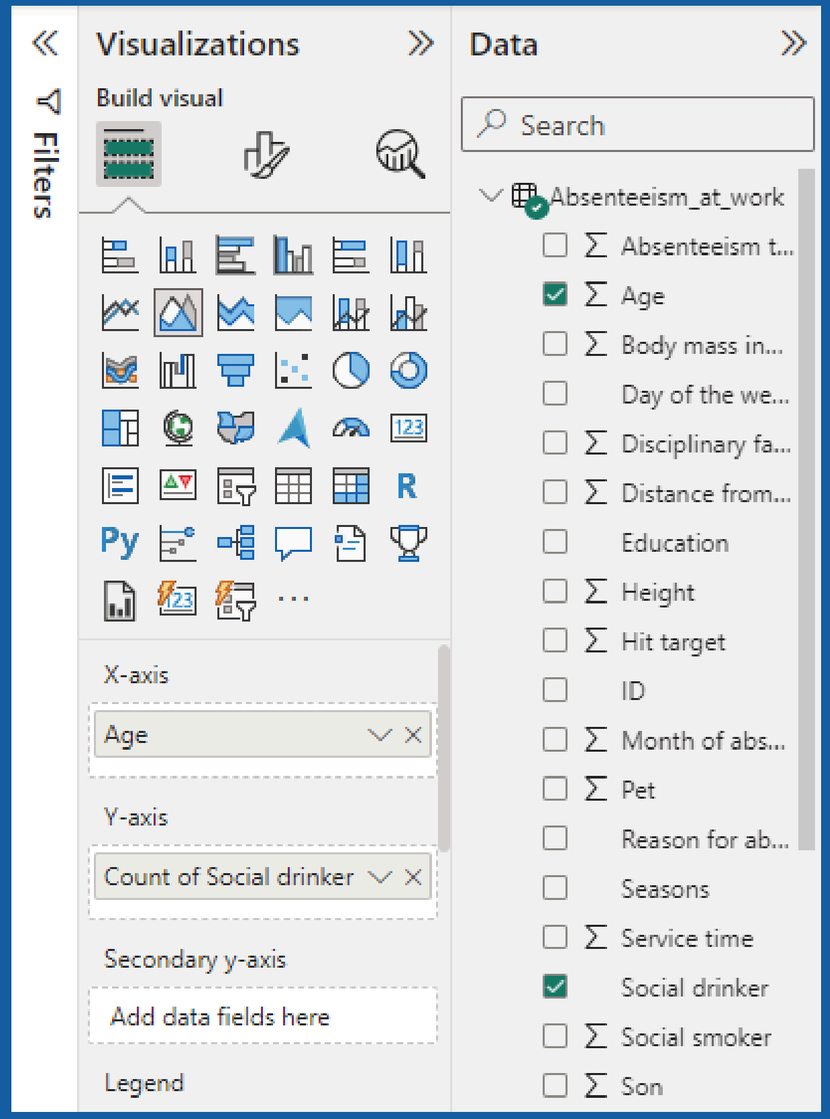
Short answer text

DATA STORAGE

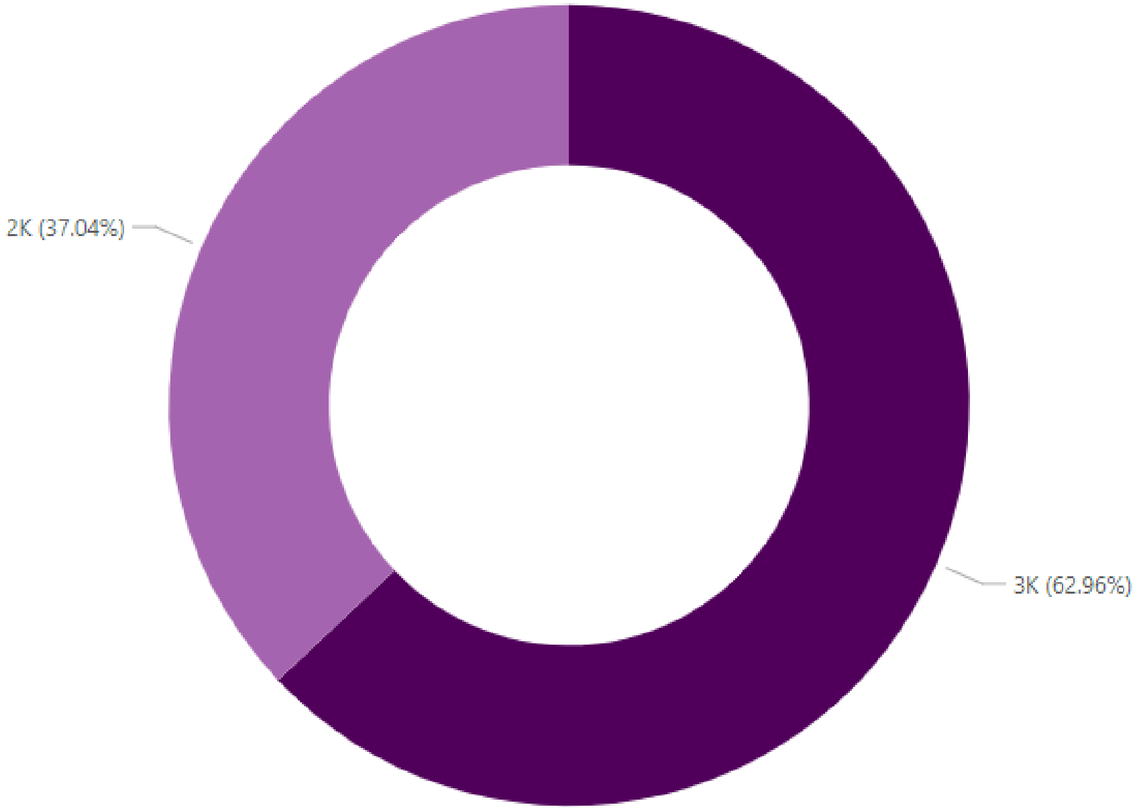


ID	Reason for absence	Age	Gender	Day of the week	Month of absence	Season
20	23	47	female	Wednesday	August	winter
30	23	38	female	Friday	August	winter
33	23	37	female	Thursday	August	winter
40	23	36	female	Friday	August	winter
42	23	37	female	Monday	September	spring
48	23	40	female	Thursday	September	spring
62	23	36	female	Friday	September	spring
64	23	37	female	Tuesday	October	spring
68	23	28	female	Friday	October	spring
70	23	28	female	Wednesday	October	spring
71	23	47	female	Wednesday	October	spring
72	23	40	female	Thursday	October	spring
82	23	28	female	Wednesday	November	spring
85	23	40	female	Thursday	November	spring
89	23	28	female	Wednesday	November	spring
91	23	28	female	Wednesday	November	spring
101	23	50	female	Wednesday	December	summer
114	23	28	female	Wednesday	January	summer
138	23	27	female	Friday	February	summer
142	23	33	female	Thursday	February	summer
148	23	28	female	Tuesday	February	summer
153	23	28	female	Tuesday	February	summer
172	23	28	female	Friday	March	autumn
175	23	28	female	Wednesday	March	autumn
186	23	28	female	Wednesday	April	autumn
189	23	34	female	Thursday	April	autumn
193	23	50	female	Thursday	April	autumn
194	23	40	female	Friday	April	autumn
224	23	40	female	Thursday	June	winter

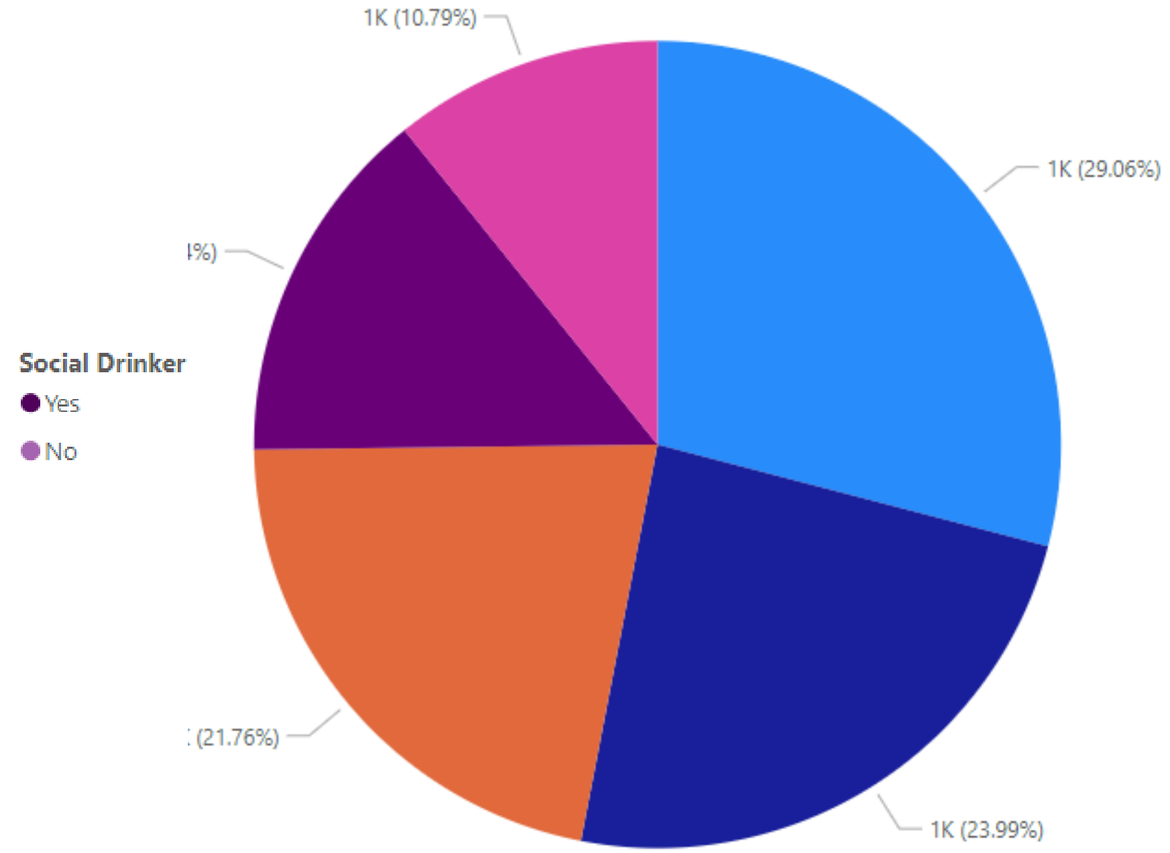

SQLQuery4.sql - LAPTOP-V9ED2IOU.Absenteeism (LAPTOP-V9ED2IOU\Saphulelo Ngcobo (70)) - Microsoft SQL Server Management Studio
Quick Launch (Ctrl+Q)
File Edit View Query Project Tools Window Help
New Query
Absenteeism
Execute
Object Explorer
Connect
LAPTOP-V9ED2IOU (SQL Server 16.0.1000.6
Databases
System Databases
Database Snapshots
Surveys
Absenteeism
Database Diagrams
Tables
System Tables
FileTables
External Tables
Graph Tables
dbo.Absenteeism_at_work
dbo.compensation
dbo.Reasons
Dropped Ledger Tables
Views
External Resources
Synonyms
Programmability
Query Store
Service Broker
Storage
Security
Security
Server Objects
Replication
Always On High Availability
Management
Integration Services Catalogs
SQL Server Agent (Agent XPs disabled)
XEvent Profiler
SQLQuery4.sql - LA...ulelo Ngcobo (70))
SQLQuery3.sql - LA...ulelo Ngcobo (68))
SQLQuery2.sql - LA...ulelo Ngcobo (66))
SQLQuery1.sql - LA...ulelo Ngcobo (63))
--Joining Tables
SELECT * FROM Absenteeism_at_work a
LEFT JOIN compensation b
ON a.ID = b.ID
LEFT JOIN Reasons r
ON a.Reason_for_absence = r.Number;
--Find the healthiest employees/Filtering through table
SELECT * FROM Absenteeism_at_work
WHERE Social_drinker = 0 and Social_smoker = 0
and Body_mass_index > 18 and Body_mass_index < 25 and Absenteeism_time_in_hours < (SELECT AVG(Absenteeism_time_in_hours) FROM Absenteeism_at_work)
--Optimize query
SELECT
a.ID,
r.Reason,
Month_of_absence,
Body_mass_index,
CASE WHEN Month_of_absence IN (9,10,11) THEN 'Spring'
WHEN Month_of_absence IN (12,1,2) THEN 'Summer'
WHEN Month_of_absence IN (3,4,5) THEN 'Autumn'
WHEN Month_of_absence IN (6,7,8) THEN 'Winter'
ELSE 'Unknown' END as Season_Names,
CASE WHEN Body_mass_index < 18.5 THEN 'Underweight'
WHEN Body_mass_index between 18.5 and 25 THEN 'Healthy'
WHEN Body_mass_index between 25 and 30 THEN 'Overweight'
WHEN Body_mass_index > 30 THEN 'Obesity'
ELSE 'Unknown' END as BMI_Category,
Month_of_absence,
Day_of_the_week,
Transportation_expense,
Social_drinker,
Social_smoker,
Disciplinary_failure,
100 %
Results
Messages
ID Reason Month of absence Body mass index Season Names BMI Category Month of absence Day of the week Transportation expense Social drinker Social smoker Disciplinary failure Age Work
Query executed successfully.
LAPTOP-V9ED2IOU (16.0 RTM)
LAPTOP-V9ED2IOU\Saphul...
Absenteeism
00:00:00
740 rows
Ready
Ln 26
Col 62
Ch 59
INS
70°F
Mostly sunny
Search
3:34 PM
4/24/2024


Filters
Visualizations
Build visual
X-axis
Age
Y-axis
Count of Social drinker
Secondary y-axis
Add data fields here
Legend
Data
Search
Absenteeism_at_work
Absenteeism t...
Age
Body mass in...
Day of the we...
Disciplinary fa...
Distance from...
Education
Height
Hit target
ID
Month of abs...
Pet
Reason for ab...
Seasons
Service time
Social drinker
Social smoker
Son

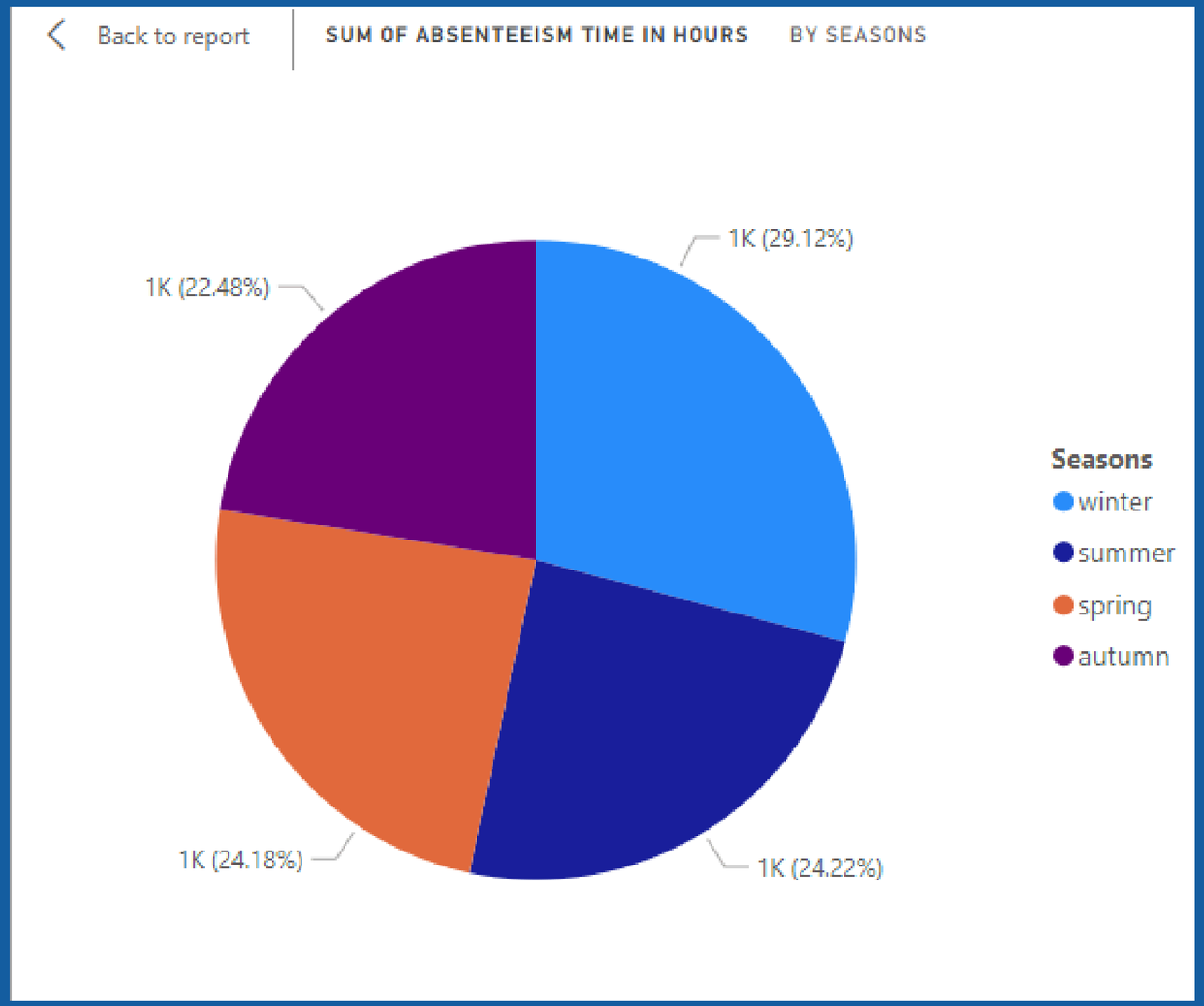
Back to report | **SUM OF DAYS ABSENT** BY SOCIAL DRINKER

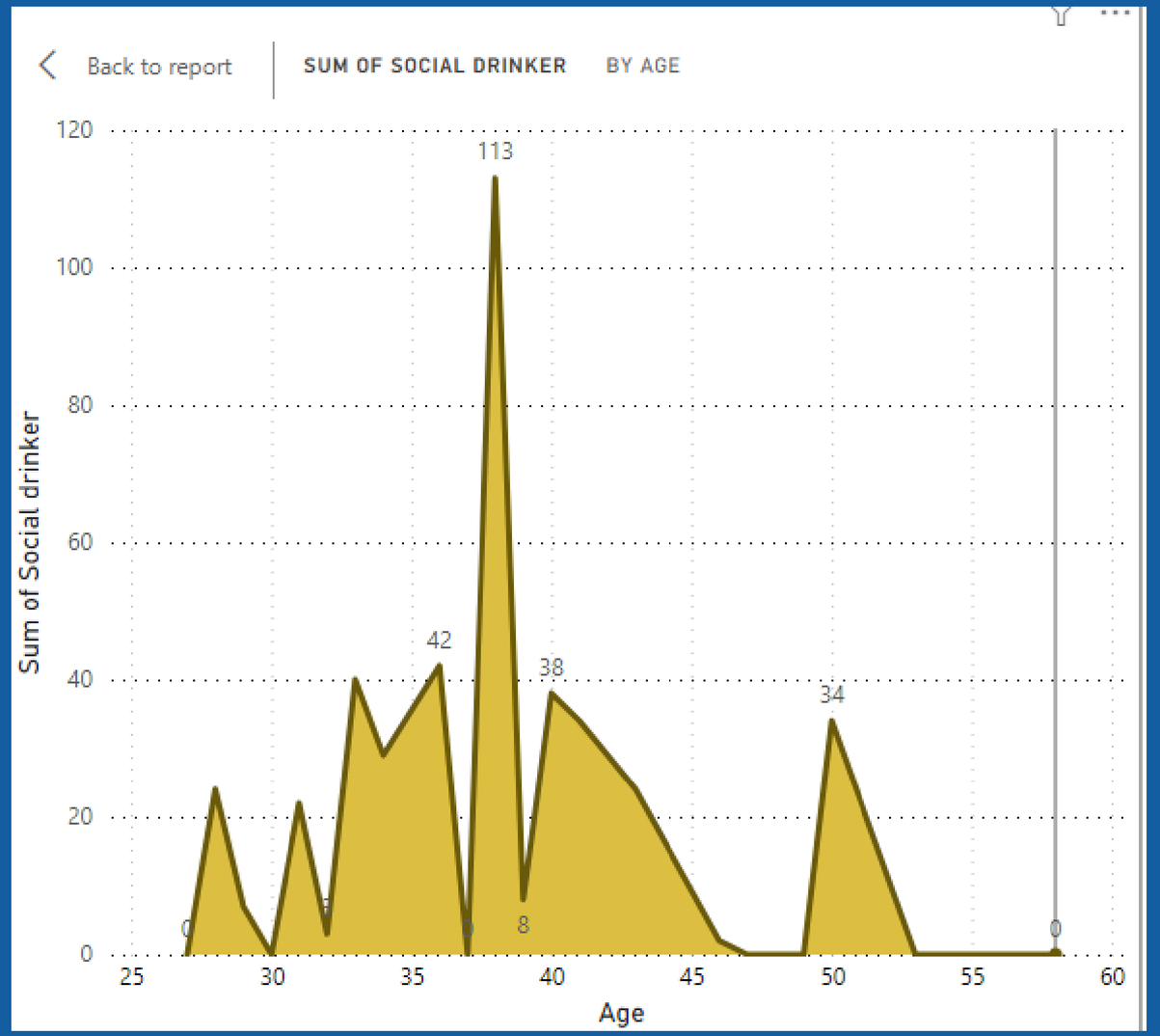


ort | **SUM OF DAYS ABSENT** BY DAY OF THE WEEK



DATA INSIGHTS





THE END

THANK YOU