



BÀI TẬP

SQL FOR DATA ANALYTICS

Lê Văn Hạnh

levanhhanhvn@gmail.com

2023

MỤC LỤC

GIỚI THIỆU CÁC CƠ SỞ DỮ LIỆU DÙNG TRONG TÀI LIỆU 1

1 BÀI TẬP SỐ 1 (*Data Definition Language*) 5

2 BÀI TẬP SỐ 2 (*Single Table*) 16

3 BÀI TẬP SỐ 3 (*Multiple Table*) 23

4 BÀI TẬP SỐ 4 (*Common Table Expression - CTE*) 28

5 BÀI TẬP SỐ 5 (*Window Functions*) 33

6 BÀI TẬP SỐ 6 (*FROM - Pivot & UnPivot*) 40

7 BÀI TẬP SỐ 7 (*GROUP BY - RollUp*) 47

8 BÀI TẬP SỐ 8 (*Optimized Queries*) 50

9 BÀI TẬP SỐ 9 (*Struct - Array & Record*) 52

10 BÀI TẬP TỔNG HỢP 55

GIỚI THIỆU CÁC CƠ SỞ DỮ LIỆU DÙNG TRONG TÀI LIỆU

i. Athletes

- Dữ liệu của table này được thu thập vào năm 2016. Table gồm 11.538 hàng (records) cho biết thông tin về các vận động viên (VDV) và thành tích đạt được của họ trong năm 2016.
- Các cột có trong table
 - **id** : mã số của VDV. Mã này là duy nhất cho từng VDV.
 - **name** : tên của vận động viên. Chưa ghi nhận có bất kỳ VDV nào trong cùng quốc gia có tên trùng nhau.
 - **nationality** : quốc tịch
 - **sex** : giới tính gồm 2 phái nữ (female) và nam(male)
 - **date_of_birth** : ngày sinh của vận động viên
 - **height** : chiều cao của vận động viên tính theo đơn vị mét.
 - **weight** : cân nặng của vận động viên tính theo đơn vị kg.
 - **sport** : gồm 28 môn: aquatics, archery, athletics, badminton, basketball, boxing, canoe, cycling, equestrian, fencing, football, golf, gymnastics, handball, hockey, judo, modern pentathlon, rowing, rugby sevens, sailing, shooting, table tennis, taekwondo, tennis, triathlon, volleyball, weightlifting, wrestling.
 - **gold** : số lượng huy chương vàng của vận động viên
 - **silver** : số lượng huy chương bạc của vận động viên
 - **bronze** : số lượng huy chương đồng của VDV
 - **total** : tổng số huy chương của VDV (= vàng + bạc + đồng)

ii. Data Science Salaries¹

Cơ sở dữ liệu gồm các table sau:

(i).- **CompanySizes**: gồm 2 field

- **Size**: là mã dùng để phân biệt loại của Công ty (khóa chính). Việc phân loại được dựa trên số người trung bình làm việc cho công ty trong năm. Gồm 3 mức L, S, M.
- **SizeName**: tên của loại Công ty.
- **NumberOfEmployee**: quy ước số lượng nhân viên theo loại Công ty.

¹ <https://www.kaggle.com/datasets/arnabchaki/data-science-salaries-2023>

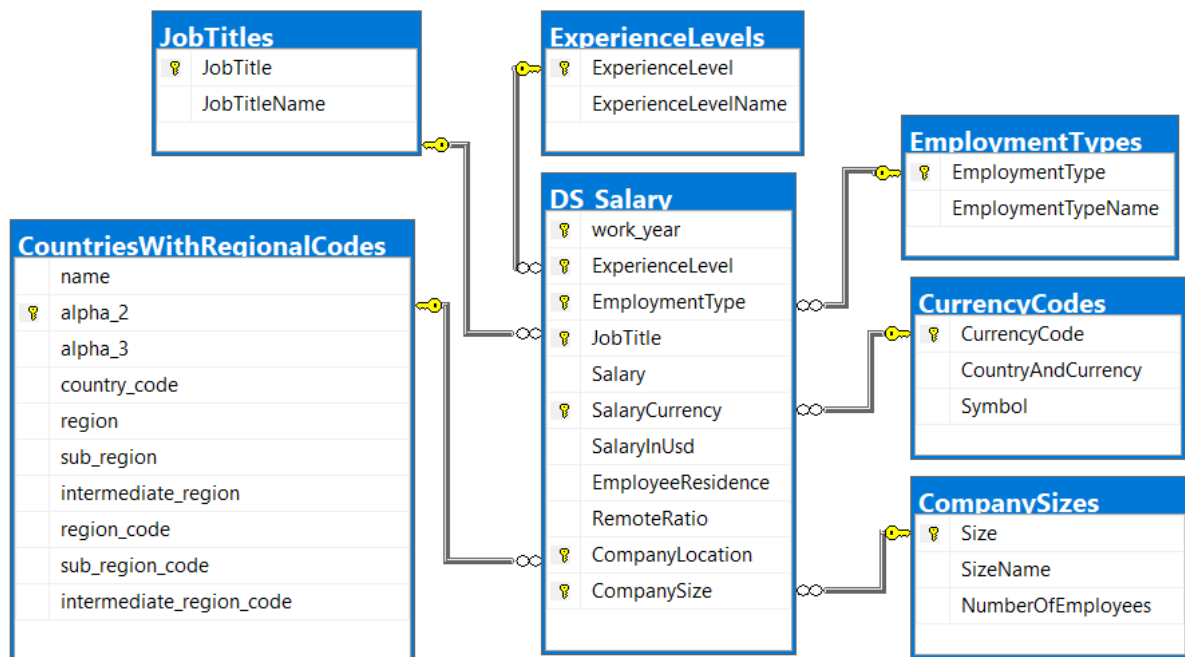
- (ii).- **CurrencyCodes**: gồm 2 field
- *CurrencyCode*: Mã đơn vị tiền tệ được dùng để trả lương cho nhân viên. Là khóa chính của table.
 - *CountryAndCurrency*: tên quốc gia sử dụng và tên gọi của đơn vị tiền tệ.
 - *Symbol*: ký hiệu dùng cho đơn vị tiền tệ
- (iii).- **EmploymentTypes**: gồm 2 field
- *EmploymentType*: mã loại nhân viên. Là khóa chính của table.
 - *EmploymentTypeName*: tên loại nhân viên
- (iv).- **Experiences**: gồm 2 field
- *ExperienceLevel*: mã cấp độ kinh nghiệm trong công việc. Là khóa chính của table.
 - *ExperienceLevelName*: tên cấp độ kinh nghiệm trong công việc.
- (v).- **JobTitles**: gồm 2 field
- *JobTitle*: mã vai trò công việc. Là khóa chính của table.
 - *JobTitleName*: tên vai trò công việc.
- (vi).- **CountriesWithRegionalCodes**: gồm 11 field
- *Name* : tên quốc gia/ vùng lãnh thổ. Tên này không trùng nhau.
 - *Alpha_2* : mã quốc gia với 2 ký tự. Mã này không trùng nhau. Được chọn làm khóa chính của table.
 - *Alpha_3* : mã quốc gia với 3 ký tự. Mã này không trùng nhau.
 - *country-code*: mã số của quốc gia. Mã này không trùng nhau.
 - *iso_3166_2*
 - *region* : tên châu lục
 - *sub_region* : tên khu vực (tiểu vùng) thuộc châu lục
 - *intermediate_region*
 - *region_code*: mã của châu lục
 - *sub-region_code*: mã số của khu vực (tiểu vùng) thuộc châu lục
 - *intermediate_region_code*
- (vii).- **DS_Salary**: gồm 11 field. Mỗi record là thông tin khảo sát từ 1 nhân viên theo năm. Table không chứa thông tin riêng của từng nhân viên nên dữ liệu trong table này có 1171 record trùng nhau (ngẫu nhiên).
- *work_year*: Năm trả lương.
 - *ExperienceLevel*: mã cấp độ kinh nghiệm trong công việc đã làm trong năm. Tương ứng field *ExperienceLevel* trong table *Experience*.
 - *EmploymentType*: mã loại nhân viên đã làm trong năm. Tương ứng field *EmploymentType* trong table *EmploymentType*.
 - *JobTitle*: mã vai trò công việc đã làm trong năm.

- *Salary*: Tổng số tiền lương gộp được trả.
- *SalaryCurrency*: mã đơn vị tiền tệ dùng để trả lương. Tương ứng field *CurrencyCode* trong table *CurrencyCodes*.
- *SalaryInUsd*: Tiền lương quy đổi ra đơn vị tiền tệ USD.
- *EmployeeResidence*: Quốc gia cư trú chính của nhân viên trong năm làm việc dưới dạng mã quốc gia. Tương ứng field *alpha-2* trong table *CountriesWithRegionalCodes*.
- *RemoteRatio*: Tỷ lệ công việc được thực hiện từ xa (làm việc online).
- *CompanyLocation*: Quốc gia của văn phòng chính hoặc chi nhánh hợp đồng của chủ lao động. Tương ứng field *alpha-2* trong table *CountriesWithRegionalCodes*.
- *Size*: là mã cho biết số người trung bình làm việc cho công ty trong năm. Tương ứng field *CompanySize* trong table *CompanySize*.

Có thể chọn 1 trong 2 khóa chính sau (mỗi khóa chính cũng đều gồm 7 fields:

- work_year, ExperienceLevel, EmploymentType, JobTitle, SalaryCurrency, Size, CompanyLocation*
- work_year, ExperienceLevel, EmploymentType, JobTitle, SalaryCurrency, Size, EmployeeResidence*

Gợi ý mối quan hệ giữa các table



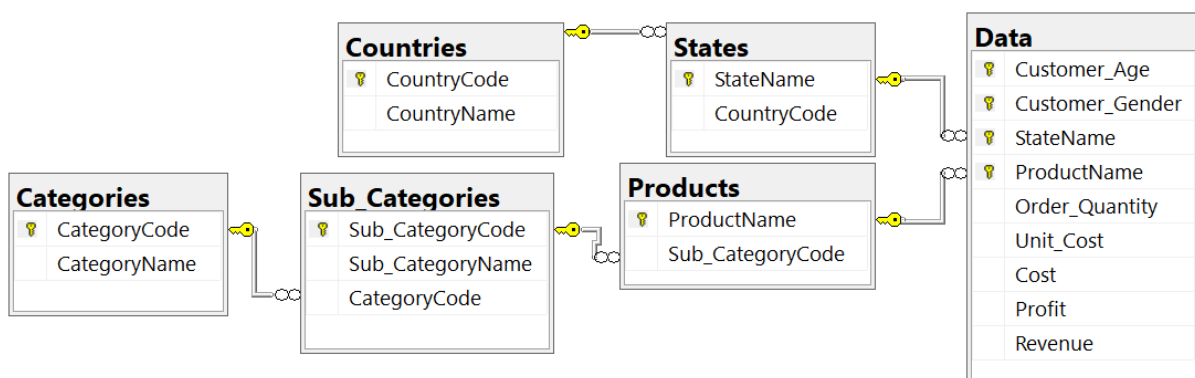
Trong diagram trên, đã chọn khóa chính là dạng (i):

iii. Sales

Cho dữ liệu về tình hình bán hàng của 1 doanh nghiệp. Các table có trong cơ sở dữ liệu gồm:

- (i).- **States**
 - *StateName*: tên của tiểu bang/tỉnh/thành phố. Khóa chính.
 - *CountryCode*: mã của quốc gia mà tiểu bang/tỉnh/thành phố thuộc về.
- (ii).- **Categories** (loại hàng)
 - *CategoryCode*: mã loại hàng. Khóa chính.
 - *CategoryName*: tên loại hàng
- (iii).- **Countries**
 - *CountryCode*: mã quốc gia. Khóa chính.
 - *CountryName*: tên quốc gia
- (iv).- **Sub_Categories** (nhóm hàng - 1 loại hàng hàng gồm nhiều nhóm hàng)
 - *Sub_CategoryCode*: mã nhóm hàng. Khóa chính.
 - *Sub_CategoryName*: tên nhóm hàng.
 - *CategoryCode*: mã loại hàng.
- (v).- **Products**
 - *Sub_CategoryCode*: Mã nhóm hàng.
 - *ProductName*: tên sản phẩm. Khóa chính.
- (vi).- **Data** (thông tin về khách hàng và những sản phẩm đã mua)
 - *Customer_Age*: tuổi của khách hàng.
 - *Customer_Gender*: giới tính của khách hàng. gồm F và M.
 - *StateName*: mã của tiểu bang/tỉnh/thành phố.
 - *ProductName*: tên sản phẩm
 - *Order_Quantity*: số lượng hàng đặt
 - *Unit_Cost*: giá vốn. Với $Cost = Order_Quantity * Unit_Cost$
 - *Cost*: trị giá.
 - *Profit*: lợi nhuận
 - *Revenue*: doanh thu. Với $Revenue = Profit + Cost$.

Với khóa chính gồm các field: *Customer_Age*, *Customer_Gender*, *StateName*, *ProductName*.



BÀI TẬP SỐ 1

(Data Definition Language)

1.1. Yêu cầu chung

- Vẽ database diagram của từng CSDL
- Xác định thứ tự tạo lập table và nhập dữ liệu.
- Tạo mới project theo dạng <tên học viên>_<tên khóa học>.

Ví dụ: LeVanHanh_K287

- Tạo mới dataset cho mỗi bài tập sau đây.
- Trong mỗi dataset sẽ viết lệnh tạo lập các table có trong phần mô tả của từng CSDL.
- Viết lệnh bổ sung khóa chính (Primary Key) và khóa ngoại (Foreign Key) cho các table.

1.2. Bài tập

1.2.1. Quản lý bán hàng

(i).- **CTHD** (Chi tiết hóa đơn)

- *Tên từ:* gồm các thuộc tính SoHD, MaHH, SoLuongHoaDon.
- *Dữ liệu minh họa:*

SoHD	MaHH	SoLuong HoaDon
HD001	H0004	6
HD001	H0027	10
HD002	H0004	15
HD002	H0005	6
HD003	H0027	5
HD004	H0006	80
HD004	H0011	70
HD004	H0022	30
HD005	H0003	9
HD006	H0006	70
HD006	H0007	15

SoHD	MaHH	SoLuong HoaDon
HD007	H0007	63
HD007	H0017	44
HD007	H0019	15
HD008	H0013	7
HD008	H0023	32
HD008	H0024	25
HD009	H0006	35
HD009	H0008	18
HD009	H0009	80
HD009	H0015	30
HD009	H0020	20
HD009	H0023	24

SoHD	MaHH	SoLuong HoaDon
HD010	H0009	10
HD010	H0014	30
HD010	H0018	24
HD011	H0019	39
HD011	H0021	70
HD012	H0001	36
HD012	H0002	12
HD012	H0010	80
HD012	H0026	60
HD013	H0012	8
HD013	H0016	65
HD013	H0025	15

(ii).- **DDH** (Đơn đặt hàng)

- *Tên từ:* gồm các thuộc tính SoDDH(*khóa chính*), NgayDatHang, NgayHenGiaoHang, TongTien, MaKH, MaNV. Trong đó SoDDH dùng để phân biệt các đơn hàng với nhau và có kiểu là số nguyên.
- *Dữ liệu minh họa:*

SoDDH	NgayDatHang	NgayHenGiaoHang	TongTien	MaKH	MaNV
1	2019-09-28	2019-09-30	6900000	K008	NV01
2	2019-09-28	2019-09-29	2808000	K001	NV04
3	2019-09-28	2019-09-28	622436000	K009	NV09
4	2019-09-29	2019-09-30	76500	K005	NV01
5	2019-09-29	2019-10-02	303625600	K009	NV01
6	2019-10-03	2019-10-03	353120000	K003	NV03
7	2019-10-04	2019-10-04	94000000	K003	NV04
8	2019-10-05	2019-10-08	588204000	K009	NV04
9	2019-11-02	2019-11-03	7125000	K005	NV09
10	2019-11-05	2019-11-06	209979200	K003	NV08
11	2019-12-07	2019-12-09	126470000	K007	NV03
12	2019-12-08	2019-12-08	34850000	K009	NV04

(iii).- **LoaiHang** (Loại hàng hóa)

- *Tên từ:* gồm các thuộc tính MaLoai (*khóa chính*), TenLoai. Trong đó MaLoai dùng để phân biệt các loại hàng với nhau.
- *Dữ liệu minh họa:*

MaLoai	TenLoai
L1	Nước uống
L2	Bia rượu
L3	Điện tử gia dụng
L4	Vật dụng

(iv).- **NhanVien** (Nhân viên)

- *Tên từ:* gồm các thuộc tính MaNV (*khóa chính*), TenNV, DiaChi, DienThoai. Mã nhân viên không được trùng nhau.
- *Dữ liệu minh họa:*

MaNV	TenNV	DiaChi	DienThoai
NV01	Nguyen Khanh Ngoc	507 Dinh Tien Hoang	08658963
NV02	Hoang Tu Anh	908 Chuong Duong	08745612
NV03	Chau Ngoc Phuong	722 Mai Xuan Thuong	08450203
NV04	Pham Hung Minh	41 Phan xich Long	08456023
NV05	Bui Cong Sinh	14 Hung Vuong	08674036
NV06	Tran Vi Minh	111 Cong Quynh	08716027
NV07	Mai Kieu Oanh	456 Nguyen Cong Tru	08476502
NV08	Pham Hoai Tram	26 Le Hong Phong	08445020
NV09	Do Thanh Anh	7 Hoang Dieu	08745230

(v).- **CTDDH** (Chi tiết đơn đặt hàng)

- *Tên từ:* gồm các thuộc tính SoDDH, MaHH, SoLuongDatHang. *Khóa chính* gồm SoDDH và MaHH.
- *Dữ liệu minh họa:*

SoDDH	MaHH	SoLuongDatHang
1	H0004	25
1	H0027	20
2	H0004	15
2	H0005	6
3	H0006	80
3	H0011	70
3	H0022	30
4	H0003	9
5	H0013	7
5	H0023	32
5	H0024	25

SoDDH	MaHH	SoLuongDatHang
6	H0019	39
6	H0021	70
7	H0012	8
7	H0016	65
7	H0025	15
8	H0001	36
8	H0002	12
8	H0010	80
8	H0026	60
9	H0009	10
9	H0014	30
9	H0018	24

SoDDH	MaHH	SoLuongDatHang
10	H0006	35
10	H0008	18
10	H0009	80
10	H0015	30
10	H0020	20
10	H0023	24
11	H0007	63
11	H0017	44
11	H0019	15
12	H0006	70
12	H0007	15

(vi).- **KhachHang** (khách hàng)

- *Tên từ:* gồm các thuộc tính MaKH (*khóa chính*), TenKH, DiaChi, DienThoai. Trong đó MaKH được sử dụng để phân biệt các khách hàng với nhau.
- *Dữ liệu minh họa:*

MaKH	TenKH	DiaChi	DienThoai
K001	Anh Hang	120 Ha Ton Quyen	8171717
K002	Bong Hong	24 Ky Con	8800256
K003	Em Anh	6 Ky Hoa	918852258
K004	Ho Han	8 Pham Van Khoe	8430156
K005	Koko Company	90 An Duong Vuong	8250102
K006	Queen Cozinha	891 An Duong Vuong	918743051
K007	Quoc Cuong	10 Tan Da	8950203
K008	Suoi Tra	2817 Minh Phung	8356210
K009	Song Trang	187 Lao Tu	9450210
K010	Vinh Vien	45 Su Van Hanh	8654790

(vii).- **HoaDon** (Hóa đơn)

- *Tên từ:* Sau khi khách hàng đặt hàng, dựa trên đặt hàng này nhân viên sẽ lập hóa đơn cho những hàng hóa mà khách đã đặt. Table HoaDon gồm các thuộc tính SoHD, LanGiao, NgayHoaDon, TongTien, SoDDH, MaNV. Trong đó, SoHD không được trùng nhau (khóa chính). Do có thể có những đơn đặt hàng chứa những mặt hàng mà cửa hàng đã hết (chờ nhập xong mới lập hóa đơn cho lần sau) do đó cùng 1 đơn đặt hàng nhưng phải giao hàng thành nhiều lần, mỗi lần là 1 hóa đơn riêng.

- *Dữ liệu minh họa:*

SoHD	LanGiao	NgayHoaDon	TongTien	SoDDH	MaNV
HD001	1	2019-09-29	1200000	1	NV01
HD002	1	2019-09-29	2808000	2	NV04
HD003	2	2019-09-30	1680000	1	NV05
HD004	1	2019-09-28	622436000	3	NV09
HD005	1	2019-10-02	76500	4	NV01
HD006	1	2019-10-03	34850000	12	NV04
HD007	1	2019-10-04	126470000	11	NV03
HD008	1	2019-10-07	303625600	5	NV01
HD009	1	2019-10-07	209979200	10	NV08
HD010	1	2019-11-03	7125000	9	NV09
HD011	1	2019-11-05	353120000	6	NV03
HD012	1	2019-12-08	588204000	8	NV04
HD013	1	2019-12-08	94000000	7	NV04

(viii).- **HangHoa** (Hàng hóa)

- *Tên từ:* gồm các thuộc tính MaHH (khóa chính), TenHH, ThoiGianBH, DVT, DonGia, MaLoai. Thuộc tính MaHH được dùng để phân biệt các hàng hóa với nhau.

- *Dữ liệu minh họa:*

MaHH	TenHH	ThoiGianBH	DVT	DonGia	MaLoai
H0001	But chì đen	12	Cây	1000	L4
H0002	But long kim	12	Cây	14000	L4
H0003	Bang nhạc Sony	12	Cái	8500	L4
H0004	Beer Tiger	12	Thùng	180000	L2
H0005	But sạp	12	Hộp	18000	L1
H0006	Coca cola	12	Thùng	80000	L1
H0007	Cassette Sharp 100W	6	Cái	1950000	L3
H0008	Cassette Sharp 500W	12	Cái	7800000	L3
H0009	Luoi lam Gillett	12	Hộp	4500	L4
H0010	May lạnh Sony 1	12	Cái	4500000	L3
H0011	May lạnh Sony 2	12	Cái	8800000	L3
H0012	Nồi cơm điện Sony	12	Cái	2500000	L3
H0013	Nồi cơm điện Toshiba	12	Cái	2300000	L3
H0014	Nước yến	12	Thùng	120000	L1
H0015	Pepsi	12	Thùng	80000	L1
H0016	Quạt bàn	12	Cái	100000	L3
H0017	Rượu chat do	6	Chai	55000	L2
H0018	Rượu Napoleon	12	Chai	145000	L2
H0019	Rượu Champagne	12	Chai	80000	L2
H0020	Tivi Sharp 14"	12	Cái	3200000	L3
H0021	Tivi Sony 20"	12	Cái	5000000	L3
H0022	Tap 100 trang	6	Quyển	1200	L4
H0023	Tap 50 trang	12	Quyển	800	L4
H0024	Tu lạnh Sharp 200 lit	12	Cái	11500000	L3
H0025	Tu lạnh Sony 100 lit	12	Cái	4500000	L3
H0026	Video Sony	12	Cái	3800000	L4
H0027	PEPSI	12	Thùng	120000	L1
H0028	CoCa	12	Thùng	135000	L1
H0029	Mirinda	6	Thùng	100000	L1

1.2.2. Quản lý đề án

Cho 1 lược đồ CSDL Quản Lý Đề Án gồm các lược đồ quan hệ sau:

(i).- NHÂNVIÊN

- *Tên từ*

- *MaNV* : Mã số của nhân viên. Mỗi nhân viên có một mã số riêng kiểu số nguyên để phân biệt với nhau (*khóa chính*).
- *HoNV* : Họ của nhân viên.
- *TenLot* : Các chữ nằm giữa họ và tên.
- *TenNV* : Tên của nhân viên.
- *NgaySinh*: Ngày sinh của nhân viên.
- *DiaChi* : Địa chỉ liên hệ của nhân viên.
- *Phai* : Nhận 1 trong 2 giá trị là “NAM” hoặc “NU”.
- *Luong* : Lương của nhân viên
- *MaPB* : Mã số của phòng ban mà nhân viên trực thuộc.

- *Dữ liệu minh họa*

<i>MaNV</i>	<i>HoNV</i>	<i>TenLot</i>	<i>TenNV</i>	<i>NgaySinh</i>	<i>DiaChi</i>	<i>Phai</i>	<i>Luong</i>	<i>MaPB</i>
1	DINH	BA	TIEN	1/9/75	731 Tran Hung Dao, Q1, TPHCM	NAM	30000	P3
2	VO	THANH	TUNG	12/8/65	106bis Tran Hung Dao, Q1, TPHCM	NAM	40000	P3
3	TRAN	THANH	TAM	7/31/82	853 Mai Thi Luu, Q1, TPHCM	NAM	25000	P2
4	VO	MANH	HUNG	9/15/72	975 Ba Ria, vung Tau	NAM	38000	P2
5	LE	NGOC	QUYEN	10/10/87	450 Trung Vương, Hà Nội	NU	55000	P1
6	TRAN	MINH	KHANG	6/24/72	Tran Binh Trong, TPHCM	NAM	58000	P1
7	LE	THI	NHAN	6/20/71	291 Ho Van Hue, QPN, TPHCM	NU	43000	P3
8	TRAN	HONG	QUANG	3/29/89	980 Le Hong Phong, Q10, TPHCM	NAM	25000	P2
9	BUI	THUY	VU	7/19/68	332 Nguyen Thai Hoc, Q1, TPHCM	NU	25000	P1

(ii).- ĐEÁN

- *Tên từ*

- *MaDA* : Mã số của đề án. Mỗi đề án có một mã số riêng để phân biệt với nhau (*khóa chính*).
- *TenDA* : Tên của đề án..
- *DDiem_DA* : Địa điểm thực hiện đề án.
- *MaPB* : Mã số của phòng ban được giao thực hiện đề án.

- *Dữ liệu minh họa*

<i>MaDA</i>	<i>TenDA</i>	<i>DDiem_DA</i>	<i>MaPB</i>
DA01	San pham X	Vũng Tàu	P3
DA02	San pham Y	Nha Trang	P3
DA03	San pham Z	TP.HCM	P3
DA04	Tin hoc hoa	Hà Nội	P2
DA05	Cap quang	TP.HCM	P1
DA06	Dao tao	Hà Nội	P2

(iii).- ĐIADIEMPBAN

- *Tên từ*

- *MaPB* : Mã số của phòng ban (*khóa chính*).
- *DiaDiem*: Địa điểm đặt bộ phận liên lạc của phòng ban. Một phòng ban có thể có nhiều địa điểm. Nhưng các địa điểm của một phòng ban nào đó không được đặt trên cùng 1 tỉnh thành phố.

- Dữ liệu minh họa

MaPB	DiaDiem
P1	TP.HCM
P2	Hà Nội
P3	Nha Trang
P3	TP.HCM
P3	Vũng Tàu

(iv).- **PHÒNGBAN**

- *Tên từ*
 - **MaPB** : Mã số của phòng ban. Dùng để phân biệt các phòng ban với nhau (*khóa chính*).
 - **TenPB** : Tên của phòng ban. TenPB
 - **TruongPhong** : Mã số nhân viên của người trưởng phòng.
 - **NgayNhanChuc** : Ngày nhận chức.
- Dữ liệu minh họa

MaPB	TenPB	TruongPhong	NgayNhanChuc
P1	QUAN LY	5	6/19/2001
P2	DIEU HANH	8	1/1/2005
P3	NGHIEN CUU	2	5/22/2008

(v).- **PHÂN CÔNG**

- *Tên từ*
 - **MaNV** : Mã số của nhân viên tham gia đề án.
 - **MaDA** : Mã số của đề án.
 - **SoGioPC** : Cho biết thời gian của từng nhân viên *được phân công* tham gia trong từng đề án.
 - **SoGioDaThamGia**: Cho biết thời gian của từng nhân viên *đã tham gia* trong từng đề án.
- Khóa chính gồm 2 field MaNV và MaDA.
- Dữ liệu minh họa

MaNV	MaDA	SoGioPC	SoGioDaThamGia
1	DA01	32.5	5
1	DA02	17.5	20
1	DA03	110	0
2	DA03	60	30
2	DA04	120	15
2	DA05	20	5
3	DA01	120	7
3	DA02	60	42

(tiếp theo)

MANV	MADA	SoGioPC	SoGioDaThamGia
4	DA03	190	5
5	DA04	140	0
5	DA05	30	32
7	DA01	170	20
8	DA04	160	85
8	DA06	20	0
9	DA04	155	26
9	DA06	35	5

(vi).- **THÂN NHÂN**

- *Tên từ*
 - **MaNV** : Mã số của nhân viên.
 - **TenTN** : Tên thân nhân của nhân viên có mã số tương ứng. Giả sử các thân nhân của cùng 1 nhân viên không bao giờ có tên (TenTN) trùng nhau
 - **Phai** : Cho biết giới tính của thân nhân.
 - **NgaySinh** : Ngày sinh của thân nhân.

- *QuanHe* : Cho biết mối quan hệ đối với nhân viên có mã số tương ứng.
- Khóa chính gồm 2 field *MaNV* và *TenTN*.
- *Dữ liệu minh họa*

<i>MaNV</i>	<i>TenTN</i>	<i>Phai</i>	<i>NgaySinh</i>	<i>QuanHe</i>
1	Chau	NU	12/31/98	con gái
1	Duy	NAM	1/1/28	Cha me
1	Phuong	NU	5/5/67	vo chong
2	Chau	NU	5/3/58	vo chong
2	Khang	NAM	10/25/63	con trai
2	Quang	NU	4/5/66	con gái
7	Dang	NAM	2/19/52	vo chong

1.2.3. Địa lý Việt Nam

Cho lược đồ CSDL Địa Lý VN gồm các lược đồ quan hệ sau:

(i).- **KhuVuc (khu vực)**

- *Tên từ*:
 - *MaKV* : là mã của khu vực, dùng để phân biệt giữa các khu vực (khóa chính). Mã này chỉ gồm 1 ký tự và thuộc 1 trong các ký tự từ A đến H.
 - *TenKV* : Tên khu vực. Tên dài không quá 25 ký tự và
- *Dữ liệu minh họa*

<i>MaKV</i>	<i>TenKV</i>	
A	Bắc Trung Bộ	
B	Đông Bắc Bộ	
C	Đồng bằng sông Cửu Long	
D	Đồng bằng sông Hồng	
E	Đông Nam Bộ	
F	Duyên hải Nam Trung Bộ	
G	Tây Bắc Bộ	
H	Tây Nguyên	

(iii).- **Tỉnh_TP (Tỉnh - Thành phố)**

- **Tên từ:**

- **MaTTP** : là mã của Tỉnh – Thành Phố. Mã này không được trùng nhau (khóa chính).
- **TenTTP** : tên đầy đủ của Tỉnh - Thành Phố.
- **MaKV** : cho biết tỉnh/thành phố thuộc khu vực nào.
- **DanSo** : dân số của tỉnh (đơn vị tính là người).
- **DienTich** : diện tích của Tỉnh - Thành Phố, kiểu dữ liệu là số thực (đơn vị tính là Km²).



- **Dữ liệu minh họa**

MaTTP	TenTTP	MaKV	DanSo	DienTich
089	An Giang	C	1908352	3536.7
077	Bà Rịa-Vũng Tàu	E	1148313	1980.8
095	Bạc Liêu	C	907236	2669
024	Bắc Giang	B	1803950	3851.4
006	Bắc Kạn	B	313905	4860
027	Bắc Ninh	D	1368840	822.7
083	Bến Tre	C	1288463	2394.6
074	Bình Dương	E	2426561	2694.7
052	Bình Định	F	1486918	6066.2
070	Bình Phước	E	994679	6877
060	Bình Thuận	F	1230808	7812.8
096	Cà Mau	C	1194476	5294.8
004	Cao Bằng	B	530341	6700.3
092	Cần Thơ	C	1235171	1439.2
048	Đà Nẵng	F	1134310	1284.9
066	Đắk Lắk	H	1869322	13030.5
067	Đắk Nông	H	622168	6509.3
011	Điện Biên	G	598856	9541
075	Đồng Nai	E	3097107	5905.7
087	Đồng Tháp	C	1599504	3383.8
064	Gia Lai	H	1513847	15510.8
002	Hà Giang	B	854679	7929.5
035	Hà Nam	D	852800	860.9
001	Hà Nội	D	8053663	3358.9
042	Hà Tĩnh	A	1288866	5990.7
030	Hải Dương	D	1892254	1668.2
031	Hải Phòng	D	2028514	1522.5

(tiếp theo)

MaTTP	TenTTP	MaKV	DanSo	DienTich
056	Khánh Hòa	F	1231107	5137.8
091	Kiên Giang	C	1723067	6348.8
062	Kon Tum	H	540438	9674.2
012	Lai Châu	G	460196	9068.8
020	Lạng Sơn	B	781655	8310.2
010	Lào Cai	G	730420	6364
068	Lâm Đồng	H	1296606	9783.2
080	Long An	C	1688547	4490.2
036	Nam Định	D	1780393	1668
040	Nghệ An	A	3327791	16493.7
037	Ninh Bình	D	982487	1387
058	Ninh Thuận	F	590467	3355.3
025	Phú Thọ	B	1463726	3534.6
054	Phú Yên	F	961152	5023.4
044	Quảng Bình	A	895430	8065.3
049	Quảng Nam	F	1495812	10574.7
051	Quảng Ngãi	F	1231697	5135.2
022	Quảng Ninh	B	1320324	6177.7
045	Quảng Trị	A	632375	4739.8
094	Sóc Trăng	C	1199653	3311.8
014	Sơn La	G	1248415	14123.5
072	Tây Ninh	E	1169165	4041.4
034	Thái Bình	D	1860447	1570.5
019	Thái Nguyên	B	1286751	3536.4
038	Thanh Hóa	A	3640128	11114.7
046	Thừa Thiên Huế	A	1128620	5048.2
082	Tiền Giang	C	1764185	2510.5

093	Hậu Giang	C	733017	1621.8
017	Hòa Bình	G	854131	4591
079	Hồ Chí Minh	E	8993082	2061
033	Hưng Yên	D	1252731	930.2

084	Trà Vinh	C	1009168	2358.2
008	Tuyên Quang	B	784811	5867.9
086	Vĩnh Long	C	1022791	1475
026	Vĩnh Phúc	D	1154154	1235.2
015	Yên Bái	G	821030	6887.7

(iv).- **GiápRanh (Giáp ranh)**

- *Tên từ:* □ *MaTTP* : là mã của tỉnh – thành phố.
- *MaBN* : là mã của nước có biên giới với Việt Nam hoặc mã của vùng biển giáp ranh với đất liền của Việt Nam.
- *Khóa chính gồm MaTTP và MaBN.*
- *Dữ liệu minh họa*

<i>MaTTP</i>	<i>MaBN</i>
022	SEA
031	SEA
034	SEA
036	SEA
037	SEA
038	SEA
040	SEA
042	SEA
044	SEA
045	SEA
046	SEA
048	SEA
049	SEA

(tiếp theo)

<i>MaTTP</i>	<i>MaBN</i>
051	SEA
052	SEA
054	SEA
056	SEA
058	SEA
060	SEA
077	SEA
079	SEA
082	SEA
083	SEA
084	SEA
094	SEA
095	SEA
096	SEA

(tiếp theo)

<i>MaTTP</i>	<i>MaBN</i>
091	SEA
011	CHN
012	CHN
010	CHN
002	CHN
004	CHN
020	CHN
022	CHN
011	LAO
014	LAO
038	LAO
040	LAO
042	LAO
044	LAO

(tiếp theo)

<i>MaTTP</i>	<i>MaBN</i>
045	LAO
046	LAO
049	LAO
062	LAO
062	CPC
064	CPC
066	CPC
067	CPC
070	CPC
072	CPC
080	CPC
087	CPC
089	CPC
091	CPC

(v).- **LangGieng (Láng giềng)**

- *Tên từ:* □ *MaTTP* : là mã của tỉnh – thành phố.
- *MaLG* : là mã của tỉnh – thành phố giáp ranh với tỉnh có mã là *MaTTP*.
- *Khóa chính gồm MaTTP và MaLG.*
- *Dữ liệu minh họa*

<i>MaTTP</i>	<i>MaLG</i>
089	087
089	091
089	092
077	075
077	079
095	093
095	094
095	096
095	091
024	022
024	020
024	019
024	001
024	027

(tiếp theo)

<i>MaTTP</i>	<i>MaLG</i>
070	072
070	067
060	068
060	058
060	075
060	077
096	091
004	002
004	008
004	020
092	087
092	086
092	091
092	093

(tiếp theo)

<i>MaTTP</i>	<i>MaLG</i>
035	017
001	026
001	017
001	033
042	040
042	044
030	031
030	033
030	034
031	034
093	094
093	091
093	086
017	014

(tiếp theo)

<i>MaTTP</i>	<i>MaLG</i>
025	026
025	001
025	014
025	015
025	017
025	008
054	066
054	064
044	045
049	051
049	062
051	062
051	064
022	030

024	030	048	049	079	072	022	031
006	004	066	064	033	034	045	046
006	020	066	068	056	054	094	084
006	019	066	067	056	066	094	086
006	008	067	068	056	058	014	015
027	030	011	014	056	068	072	080
027	001	011	012	062	064	019	026
027	033	075	079	012	010	019	008
083	086	075	068	012	015	019	001
083	084	087	080	012	014	038	014
083	082	087	082	020	022	038	017
074	070	087	086	020	019	038	037
074	079	002	015	010	015	046	049
074	075	002	010	080	079	046	048
074	072	002	008	080	082	082	079
052	051	035	001	036	034	082	086
052	054	035	033	036	037	084	086
052	064	035	034	040	038	008	015
070	068	035	036	037	017	008	026
070	075	035	037	058	068	015	008

(vi).- **Bien_Nuoc (Biển – Nước)**

- **Tên từ:**

- **MaBN** : là mã của nước có biên giới với Việt Nam hoặc mã của vùng biển giáp ranh với đất liền của Việt Nam (khóa chính).
- **Ten** : tên đầy đủ của nước hoặc vùng biển giáp ranh với Việt Nam.

- **Dữ liệu minh họa**

MaBN	Tên
CPC	Campuchia
LAO	Lào
CHN	Trung quốc
SEA	Biển Đông + Vịnh Thái Lan

1.2.4. Tour du lịch

Cho lược đồ CSDL sau để quản lý các tour du lịch:

(i).- **DIEMTQ**(MaDTQ, TenDTQ, Noidung, Ynghia)

Tân từ: Mỗi điểm tham quan (DIEMTQ) có mã số điểm tham quan (MaDTQ) duy nhất để phân biệt, có tên gọi (TenDTQ) cùng với nội dung tham quan (Noidung) và ý nghĩa của điểm tham quan này (Ynghia).

(ii).- **TOUR**(MaTour, TenTour, Songay)

Tân từ: Công ty có nhiều tour du lịch, mỗi tour có mã số duy nhất (MaTour) để phân biệt cùng với tên tour (TenTour), số ngày (Songay) tham gia tour.

(iii).- **CT_THAMQUAN**(MaTour, MaDTQ, Thoigian)

Tân từ: Mỗi tour du lịch sẽ đi qua nhiều điểm tham quan, tại mỗi điểm tham quan sẽ ở lại bao nhiêu ngày (Thoigian).

(iv).- **HOPDONG**(SoHD, NgaylapHD, NoidungHD, SoNguoidi, Noidon, NgaydiHD, MaTour, MaDoan)

Tân từ: Đây là hợp đồng dành cho khách du lịch đi theo đoàn, mỗi hợp đồng có số hợp đồng (SoHD) là duy nhất để phân biệt cùng với các thông tin: ngày lập hợp đồng (NgaylapHD), nội dung của hợp đồng (NoidungHD), số người đi trong đoàn (SoNguoidi), nơi công ty sẽ đón (Noidon), ngày bắt đầu xuất phát (NgaydiHD). Hợp đồng cũng cho biết đoàn (MaDoan) sẽ đi tour nào (MaTour) tương ứng.

(v).- **DOAN**(MaDoan, TenDoan, HoTenNDD, Diachi, Dienthoai)

Tân từ: Đây là thông tin của khách hàng đại diện cho nhóm khách đi du lịch theo đoàn (DOAN). Đoàn có thể là công ty, cơ quan, xí nghiệp, trường học hay gia đình... Mỗi đoàn có một mã đoàn duy nhất (MaDoan) để phân biệt, có tên đoàn (TenDoan), họ tên người đại diện (HoTenNDD) cùng với địa chỉ liên lạc (Diachi) và điện thoại liên hệ (Dienthoai). Nếu đoàn khách không phải là công ty thì tên đoàn là NULL.

(vi).- **DIEMDUNGCHAN**(MaDDC, TenDDC, Thanhpho)

Tân từ: Mỗi tour du lịch sẽ đi qua nhiều điểm dừng chân (MaDDC). Ứng với mỗi điểm dừng chân sẽ có một tên gọi (TenDDC) và thuộc một thành phố tương ứng (Thanhpho).

(vii).- **LOTRINH**(MaNoiDi, MaNoiDen)

Tân từ: Lộ trình (LOTRINH) mô tả chặng đường từ điểm dừng chân này (MaNoiDi) đến điểm dừng chân khác (MaNoiDen).

(viii).- **LOTRINH_TOUR**(MaTour, MaNoiDi, MaNoiDen, SongayO, SongaydicuaPT, LoaiPhuongtien, LoaiKhachsan)

Tân từ: Mỗi tour du lịch sẽ đi qua nhiều lộ trình, tại mỗi lộ trình sẽ có số ngày ở (SongayO) tương ứng, số ngày di chuyển của phương tiện (SongaydicuaPT), loại

phương tiện sử dụng trong lộ trình (như Máy bay, ô tô, tàu hỏa, tàu cao tốc...) và loại khách sạn sẽ ở (ví dụ 2 sao, 3 sao, ...).

(ix).- **CHUYEN**(MaChuyen, Tenchuyen, NgaydiCuaChuyen, MaNVHDDL, MaTour)

Tân từ: Chuyến (CHUYEN) là tour du lịch dành cho các khách lẻ đăng ký, không đi theo đoàn. Khách có thể đăng ký tại các đại lý. Thông tin của chuyến bao gồm: tên chuyến (Tenchuyen), ngày đi (NgaydiCuaChuyen), nhân viên sẽ hướng dẫn du lịch (MaNVHDDL), tour khách sẽ đi (MaTour).

(x).- **HOPDONG_NV**(SoHD, MaNVHDDL, NoidungHD_NV)

Tân từ: Khi khách hàng đi du lịch theo đoàn, công ty sẽ phân công nhân viên hướng dẫn đi cùng, một hợp đồng sẽ được lập (HOPDONG_NV). Nội dung của hợp đồng bao gồm: số hợp đồng (SoHD), mã nhân viên (MaNVHDDL) (cặp đôi thuộc tính này là duy nhất để phân biệt) cùng với nội dung của hợp đồng.

(xi).- **NHANVIENHDDL**(MaNVHDDL, TenNV, NgaysinhNV, PhaiNV, DiachiNV, DienthoaiNV)

Tân từ: Công ty có nhiều nhân viên hướng dẫn du lịch (NHANVIENHDDL), mỗi nhân viên có mã số để phân biệt (MaNVHDDL), tên nhân viên (TenNV), ngày sinh (NgaysinhNV), phái (PhaiNV) cùng với địa chỉ liên lạc (DiachiNV) và điện thoại liên hệ (DienthoaiNV).

BÀI TẬP SỐ 2 (Single Table)

2.1. Tạo project-dataset-table

1. Mở Google Cloud Console tại địa chỉ: <https://console.cloud.google.com> và đăng nhập với Google Account
2. Tạo project theo mẫu sau: <tên học viên>-BiqQuery-K287
3. Trong project mới tạo, tạo dataset tên *athletes_dataset*
4. Trong *athletes_dataset*, tạo table *athletes* từ tập tin *athletes.csv* (được GV cung cấp). Chọn table *athletes* và thực hiện các yêu cầu sau:
 - a. Xem cấu trúc table ~ SCHEMA
 - b. Xem thông tin chi tiết ~ DETAILS
 - c. Xem dữ liệu ~ PREVIEW

2.2. Thực hiện các truy vấn sau

2.2.1. Truy vấn đơn giản

1. Truy vấn tất cả dữ liệu có trong table.
2. Cho biết thông tin về *name*, *nationality*, *date_of_birth*, *height*, *weight*, *total*, *sport* của các VDV (VDV) nữ.
3. Cho biết *name*, *date_of_birth* của tất cả VDV, nhưng sắp xếp kết quả theo *date_of_birth* và những hàng có giá trị NULL sẽ được đưa lên đầu của kết quả truy vấn.
4. Tương tự câu trên (vẫn sắp xếp kết quả theo *date_of_birth*) nhưng cho những hàng có giá trị NULL chuyển xuống cuối của kết quả truy vấn.
5. Xem danh sách các VDV nữ không có huy chương (*total*=0).
6. Cho biết tên những nước mà thông tin *date_of_birth* của VDV bị thiếu (null).
7. Cho biết tên những nước có vận động viên nữ tham gia môn *rugby sevens* (*sport*=*rugby sevens*).
8. Cho biết tên những VDV chỉ đạt huy chương vàng nhưng không đạt huy chương bạc và đồng.
9. Cho biết tên những VDV nữ chỉ đạt huy chương vàng gold.
10. Cho biết *name*, *nationality*, *sex*, *date_of_birth* của những VDV nữ có huy chương vàng môn *cycling* khi có tuổi chưa đến 23. Nhắc lại dữ liệu trong table này thu thập về thành tích của VDV trong năm 2016.

2.2.2. SELECT * EXCEPT

11. Cho biết nội dung của tất cả các field, ngoại trừ field *total* của những VDV có bất kỳ huy chương nào khi đã trên 50 tuổi.

12. Cho biết nội dung của tất cả các field, ngoại trừ 2 field height và weight của những VDV người Marocco (MAR) bị thiếu thông tin về ngày sinh (*date_of_birth=null*).

2.2.3. *SELECT * REPLACE*

13. Giả sử số tiền thưởng cho mỗi huy chương như sau: gold=10.000, silver=5.000, bronze=3.000. Yêu cầu cho biết các thông tin trong table, trong đó thay giá trị cột total thành tổng số tiền thưởng mà VDV đó được nhận và chỉ tính cho những người có huy chương (total>0).

2.2.4. *Toán tử [NOT] LIKE - hàm Starts_with*

Mỗi câu sau đây yêu cầu HV thực hiện bằng cả 2 cách: LIKE và Starts_With (nếu được)

14. Cho biết tên VDV và tên nước của VDV đó sao cho 3 ký tự đầu của tên VDV là 'Car'. Yêu cầu thực hiện bằng 2 cách: sử dụng toán tử *LIKE* và hàm *Starts_with*.
15. Cho biết tên VDV và tên nước của VDV đó sao cho ký tự thứ ba của tên VDV là ký tự 'o' và ký tự thứ năm là 'a'.
16. Tìm tên (name) những VDV có ký tự thứ 3 không phải là chữ 'd' và ký tự thứ 5 không phải là khoảng trắng.
17. Tìm tên (name) những VDV có ký tự đầu tiên là chữ 'S' và ký tự thứ 5 không phải là 1 trong 2 ký tự 'm' hoặc 'n'.

2.2.5. *Toán tử [NOT] BETWEEN*

Thực hiện các yêu cầu này bằng 2 cách: có và không có sử dụng toán tử BETWEEN

18. Cho biết những VDV nữ nào có chiều cao từ 1,6m đến 1.8m, trọng lượng nhẹ hơn 60 kg nhưng vẫn đạt huy chương vàng.
19. Cho biết tên những VDV đạt huy chương vàng khi tuổi không nằm trong khoảng từ 18 đến 53.

2.2.6. *LIMIT*

20. Truy vấn 1000 hàng đầu trong dữ liệu.
21. Cho biết 10 VDV có nhiều huy chương nhất (tìm dựa trên field **total**)
22. Cho biết 10 VDV nữ có nhiều huy chương vàng nhất.
23. Cho biết tên VDV nữ trẻ tuổi nhất của Việt Nam có đạt huy chương (không phân biệt gold, silver hay bronze).
24. Trong số những vận động viên nữ, cho biết tên VDV của nước Mỹ có số huy chương vàng nhiều thứ 3 của nước Mỹ.

2.2.7. Sub query

↪ Lần lượt thực hiện các truy vấn sau (từ câu 25 đến câu 27) bằng 2 cách:

- Sử dụng LIMIT.
- Sử dụng sub query.

Dựa trên kết quả truy vấn để thấy được ưu điểm của mỗi cách viết truy vấn.

25. Cho biết tên 1 VĐV nữ đạt huy chương vàng nhưng “nhẹ cân” nhất trong những VĐV nữ đạt huy chương vàng.
26. Cho biết tên của nữ VĐV có cân nặng nhẹ nhất trong tất cả các VĐV người Argentina (nationality=ARG) từng đạt huy chương vàng.
27. Cho biết tên VĐV, quốc tịch của VĐV nam có trọng lượng nhẹ nhất.

2.2.8. Aggregate functions

28. Cho biết số hàng dữ liệu (records) có trong table.
29. Có bao nhiêu nước có VĐV đạt huy chương đồng (bronze>0 và không quan tâm đến môn mà VĐV thi đấu).

↪ Từ đây trở đi, trong mỗi câu của phần Agregate functions, yêu cầu học viên thực hiện bằng 2 cách:

- Sử dụng mệnh đề GROUP BY.
- Sử dụng mệnh đề OVER (nếu có thể vận dụng được).

30. Giả sử số tiền thưởng cho mỗi huy chương như sau: gold=10.000, silver=5.000, bronze=3.000. Cho biết tổng số tiền thưởng mà mỗi nước sẽ nhận được.
31. Thống kê số lượng VĐV của mỗi nước tham gia gồm: mã quốc gia, số lượng nam, số lượng nữ, tổng số VĐV. Kết quả được sắp xếp giảm dần theo số lượng VĐV nữ, nếu số lượng này bằng nhau sẽ sắp tăng dần theo số lượng VĐV nam.
32. Cho biết tên mỗi nước, tuổi nhỏ nhất, tuổi lớn nhất của những VĐV nước đó?
33. Cho biết tên nước, số lượng nam, nữ của mỗi nước theo minh họa sau:

nationality	sex	SoLuongVĐV
ZIM	female	26
ZIM	male	9
ZAM	female	2
ZAM	male	5
YEM	female	1
YEM	male	2
VIN	female	2
VIN	male	2
VIE	female	13
VIE	male	10
...

34. Cho biết tên tất cả các nước và số lượng VĐV nữ đạt được huy chương vàng. Các nước không có nữ đạt huy chương vàng vẫn hiện tên nước đó, nhưng lúc này số lượng sẽ là 0
35. Cho biết tên tất cả các nước và số lượng VĐV nữ đạt được huy chương vàng. Chỉ hiện tên các nước có nữ đạt huy chương vàng

36. Cho biết *nationality*, số lượng từng loại huy chương (*gold*, *silver* và *bronze*) của những VĐV nữ (*sex* = 'female') trong môn judo (*sport=judo*). Sắp xếp giảm dần theo từng số lượng huy chương. Nếu số lượng huy chương *gold* bằng nhau thì sắp xếp dựa trên *silver*, tương tự nếu *gold* và *silver* bằng nhau sẽ xét tiếp trên *bronze*.

37. Cho biết tên 3 nước có VĐV “nặng ký” nhất (trọng lượng lớn nhất).

↳ Lần lượt thực hiện các truy vấn sau sau (từ câu 37 đến câu 38) bằng 3 cách:

- ANY_VALUE
- MAX_BY/MIN_BY
- Không sử dụng các hàm ANY_VALUE và MAX_BY/MIN_BY.

38. Cho biết tên môn thể thao nào mà nước Mỹ có nhiều huy chương bạc nhất.

39. Cho biết tên nước có VĐV “lớn tuổi” và “nhỏ tuổi” nhất. Kết quả truy vấn:

Có sử dụng các hàm ANY_VALUE và MAX_BY/MIN_BY	Không sử dụng các hàm ANY_VALUE và MAX_BY/MIN_BY																												
<table><tr><th>max_age</th><th>min_age</th></tr><tr><td>NZL</td><td>TKM</td></tr></table>	max_age	min_age	NZL	TKM		<table><tr><th>max_age</th><th>min_age</th></tr><tr><td>null</td><td>TKM</td></tr><tr><td>null</td><td>MYA</td></tr><tr><td>null</td><td>CHN</td></tr><tr><td>null</td><td>ROU</td></tr><tr><td>null</td><td>AZE</td></tr><tr><td>null</td><td>GHA</td></tr><tr><td>null</td><td>NEP</td></tr><tr><td>null</td><td>LAO</td></tr><tr><td>NZL</td><td>null</td></tr><tr><td>AUS</td><td>null</td></tr></table>	max_age	min_age	null	TKM	null	MYA	null	CHN	null	ROU	null	AZE	null	GHA	null	NEP	null	LAO	NZL	null	AUS	null	
max_age	min_age																												
NZL	TKM																												
max_age	min_age																												
null	TKM																												
null	MYA																												
null	CHN																												
null	ROU																												
null	AZE																												
null	GHA																												
null	NEP																												
null	LAO																												
NZL	null																												
AUS	null																												

Học viên tự đưa ra nhận xét về cách viết truy vấn cùng kết quả của truy vấn.

40. Lần lượt thực hiện các thống kê sau:

Kết quả thực hiện câu 40									
	<i>nationality</i>	<i>age</i>		<i>Row</i>	<i>nationality</i>	<i>age</i>	<i>Row</i>	<i>nationality</i>	<i>age</i>
	NZL	62							
	AUS	62							
	GBR	61							
	NZL	60							
	AUS	60							
	GBR	59							
							
Câu a			Câu b			Câu c			

- Thống kê tên các quốc gia, tuổi giảm dần của các VĐV.
- Dựa vào câu truy vấn a, hiệu chỉnh lại để lấy ra 5 nước (không trùng nhau) có VĐV lớn tuổi nhất.
- Kết quả thực hiện tương tự như kết quả của câu b, nhưng yêu cầu thực hiện bằng cách sử dụng hàm ARRAY_AGG.
- Kết quả thực hiện tương tự như kết quả của câu b và c, nhưng yêu cầu thực hiện bằng cách sử dụng hàm ARRAY.

41. Tạo thống kê số lượng VDV theo từng chiều cao (height)

HeightOfAthletes	Quantity
1.2	1
1.3	3
1.4	19
1.5	167
1.6	1729
1.7	3168
1.8	3643
1.9	2207
2.0	519
2.1	79
2.2	3

42. Chỉ số BMI (Body Mass Index - chỉ số khối cơ thể) giúp mọi người tự kiểm tra sức khỏe dựa trên công thức:

BMI: cân nặng (kg) / (chiều cao (m) * chiều cao (m))

Tra chỉ số BMI vừa có trong bảng sau người ta biết được mức độ béo phì.

BMI	Đánh giá
<18.5	Thiếu cân (<i>Underweight</i>)
Từ 18.5 đến 22.99	Bình thường (<i>Normal</i>)
Từ 23 đến 24.99	Thừa cân (<i>Overweight</i>)
Từ 25 trở lên	Béo phì (<i>Obese</i>)

Yêu cầu: xác định số lượng VDV của từng mức đánh giá, nhưng chỉ tính cho những VDV môn cử tạ (*sport= weightlifting*). Kết quả truy vấn có dạng

underweight	normal	overweight	obese
1	38	56	161

2.2.9. Conditional expressions

Yêu cầu: sử dụng tất cả các biểu thức điều kiện (CASE, IF, ...) mà học viên biết để xử lý cho các câu sau đây:

43. Cho biết VDV với name= Kelsi Worrell có đạt huy chương vàng môn aquatics (*sport= aquatics*) hay không? Trả lời 'Có' hoặc 'không'.
44. Cho biết VDV 'Hoang Xuan Vinh' của Việt Nam có đạt được huy chương vàng môn bắn súng (*shooting*) hay không? Trả lời 'Có' hoặc 'không'.
45. Cho biết tỷ lệ VDV nữ có giải / số lượng VDV nam có giải của Việt Nam (VIE) có lớn hơn 40% hay không? Trả lời 'Có' hoặc 'không'.

Lưu ý: cần kiểm tra trước xem số lượng VDV nam có giải của Việt Nam có >0 hay không trước khi tính toán tỷ lệ, vì nếu số lượng này =0 sẽ dẫn đến phép chia có mẫu số =0 sẽ gây lỗi.

- 🔄 Lần lượt thực hiện các truy vấn sau sau (từ câu 46 đến câu 48) bằng 3 cách
- SUM
 - COUNT
 - COUNTIF

46. Cho biết tên nước, số lượng nam, số lượng nữ theo hình minh họa sau:

nationality	soluongnu	soluongnam
ZIM	26	9
ZAM	2	5
YEM	1	2
VIN	2	2
VIE	13	10
...

47. Cho biết số lượng VDV tham gia theo từng nhóm tuổi: dưới 14, từ 15-16, từ 17-18, từ 19-22 và trên 22 của mỗi nước. Thông tin gồm *nationality*, *LessThan14*, *From15To16*, *From17To18*, *From19To22*, *OlderThan22*. Gợi ý: dựa trên năm sinh và năm tham gia (2016).

48. Thực hiện thống kê gồm tên nước, số lượng VDV nữ đạt huy chương vàng của nước đó, tương tự cho huy chương bạc và đồng. Kết quả sắp xếp giảm dần theo số lượng VDV đạt huy chương vàng.

nationality	FemaleWithGold	FemaleWithSilver	FemaleWithBronze
USA	74	26	35
RUS	38	17	22
GER	29	14	23
CHN	28	21	22
GBR	27	18	15

2.2.10.Subquery

49. Có bao nhiêu nước trong danh sách. Yêu cầu viết bằng 2 cách: có và không có sử dụng subquery
50. Cho biết tên và số lượng huy chương vàng của những VDV mà những VDV này có số huy chương vàng nhiều hơn VDV có tên là 'Usain Bolt'.
51. Cho biết tên và chiều cao của 4 VDV có chiều cao là cao nhất trong tất cả các VDV. Nếu có nhiều VDV cùng có chiều cao như VDV thứ 4 thì lấy tất cả những người này (khi đó danh sách kết quả có thể có nhiều hơn 4 VDV).
52. Cho biết tên nước, số lượng huy chương đồng mà nước đó đạt được giống như số lượng huy chương đồng mà nước Hàn Quốc (KOR) đã đạt. Yêu cầu kết quả không có tên nước Hàn Quốc

2.2.11.Lớn/nhỏ nhất

🔗 Lần lượt thực hiện các truy vấn trong phần này bằng 2 cách:

- Sử dụng LIMIT.
- Sử dụng sub query.

Dựa trên kết quả truy vấn để thấy được ưu điểm của mỗi cách viết truy vấn.

53. Cho biết tên những nước có vận động nhỏ tuổi nhất trong tất cả các VDV tham gia.

54. Cho biết tên những VDV nước Mỹ có chiều cao là cao nhất trong số tất cả các VDV tham dự

<i>name</i>	<i>nationality</i>	<i>height</i>
Kevin Cordes	USA	2.21

55. Cho biết tên những VDV có chiều cao thấp nhất trong số những VDV đạt huy chương. Minh họa kết quả cần thực hiện:

<i>name</i>	<i>nationality</i>	<i>height</i>	<i>total</i>
Yan Wang	CHN	1.4	1

56. Cho biết tên những nước và số lượng VDV nữ của nước đó đạt được huy chương vàng nhiều nhất so với các nước khác.
57. Cho biết tên VDV có chiều cao thấp hơn chiều cao trung bình của tất cả VDV nước Mỹ.
58. Cho biết tên VDV Việt nam có chiều cao cao hơn chiều cao trung bình của tất cả VDV nước Mỹ.
59. Cho biết tên nước có nhiều huy chương vàng nhất.
60. Cho biết tên những nước đạt được nhiều huy chương vàng nhất của môn golf.
61. Cho biết môn thi đấu có nhiều VDV tham gia nhất.
62. Cho biết tên những nước tham gia nhiều môn thi đấu (sport) nhất.
63. Cho biết tên những nước có nhiều VDV không đạt huy chương nhất (tức là những VDV này có tên trong dữ liệu nhưng không đạt được bất kỳ huy chương nào trong 3 huy chương gold, silver, bronze).
64. Cho biết tên những nước mà tổng số huy chương vàng của họ lớn hơn tổng số huy chương bạc và đồng cộng lại.
65. Cho biết tên nước có số lượng VDV tham gia 2 môn judo và taekwondo là nhiều nhất.

2.2.12. Self Join

66. Cho biết tên những VDV có tháng và năm sinh trùng nhau.

BÀI TẬP SỐ 3 (Multiple Table)

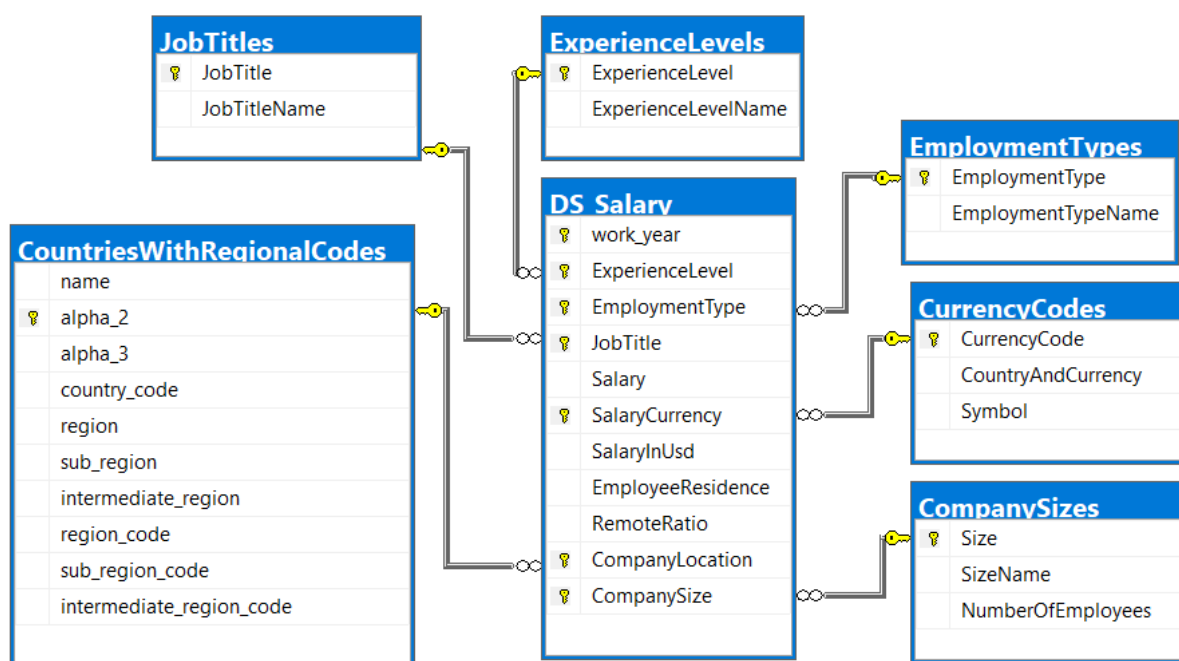
3.1. Tạo project – dataset - table

1. Mở Google Cloud Console tại địa chỉ: <https://console.cloud.google.com> và đăng nhập với Google Account
2. Tạo project theo mẫu sau: <tên học viên>-BiqQuery-K287
3. Trong project <tên học viên>-BiqQuery-K287, Thực hiện tạo dataset *DataScienceSalaries* và các table dựa trên dữ liệu do GV cung cấp.

Lưu ý: sau khi tạo table từ các file dữ liệu do GV cung cấp. Sẽ có 1 số table tên field sẽ không đúng với mô tả của dữ liệu. Học viên cần dùng lệnh để đổi tên các field này.

4. Viết lệnh bổ sung khóa chính, khóa ngoại cho các table.

Gợi ý mối quan hệ giữa các table



Như đã giới thiệu trong phần ‘Giới thiệu các CSDL dùng trong tài liệu’, có thể chọn 1 trong 2 khóa chính sau (mỗi khóa chính cũng đều gồm 7 fields):

- (i).- *work_year, ExperienceLevel, EmploymentType, JobTitle, SalaryCurrency, CompanySize, CompanyLocation*
- (ii).- *work_year, ExperienceLevel, EmploymentType, JobTitle, SalaryCurrency, CompanySize, EmployeeResidence*

Trong diagram trên, đã chọn khóa chính là dạng (i):

3.2. Thực hiện các truy vấn sau

- (1).- Cho biết mã tiền tệ (*symbol*) của nước Brazil là gì?
- (2).- Cho biết tên các Châu lục (*Region*) và các tiểu vùng của châu lục (*sub-region*) được ghi nhận trong table *CountriesWithRegionalCodes*. Bỏ qua các dòng kết quả là null (nếu có)

region	sub_region
Africa	Northern Africa
Africa	Sub-Saharan Africa
Americas	Northern America
Americas	Latin America and the Caribbean
Asia	Eastern Asia
...	...

(tổng cộng có 17 records)

- (3).- Trả lời 'có' hoặc 'không'. Cho biết tên đơn vị tiền tệ (*CountryAndCurrency*) là 'India Rupee' có được các công ty có văn phòng đặt tại 'India' dùng để trả lương cho những nhân viên thuộc loại (*EmpoymentTypeName*) 'Freelancer'.

- (4).- Cho biết những công ty trả lương (*Salary Currency*) bằng 'Singapore Dollar' thì văn phòng chính của Công ty được đặt ở nước nào (gợi ý: dựa trên *CompanyLocation*)?

name
Singapore
Finland
Israel
Câu 4

- (5).- Mức lương thấp nhất của các công ty có văn phòng đặt tại Châu Á (Asia) là bao nhiêu?
- (6).- Trong năm 2022, cho biết mức lương cao nhất, mức lương thấp nhất, chênh lệch giữa 2 mức trên của những người có tỷ lệ làm việc online (*RemoteRatio*) là 100%. Các mức lương được tính dựa trên USD (*SalaryInUsd*). Minh họa kết quả cần thực hiện:

MaxSalary	MinSalary	Difference
405000	5132	399868

- (7).- Có bao nhiêu nhân viên mà nơi thường trú của họ (*EmployeeResidence*) và nơi văn phòng công ty (*CompanyLocation*) ở 2 nước khác nhau.
- (8).- Biết rằng có 3 mức làm việc online là 0%, 50%, 100%. Thống kê tỷ lệ phần trăm các mức làm việc online trên tổng số record có trong table *DS_Salaries*.

FullOnline	HalfOnline	Offline
43.75	5.03	51.21

- (9).- Cho biết có bao nhiêu nhân viên có chức danh (*JobTitleName*) là 'Data Scientist' và đang làm việc tại nước Canada?
- (10).- Tính thu nhập bình quân theo năm của những nhân viên có chức danh (*JobTitleName*) là 'Business Intelligence Engineer'.
- (11).- Cho biết đơn vị tiền tệ (*CountryAndCurrency*) là 'India Rupee' có được các công ty dùng để trả lương cho những nhân viên thường trú (*EmployeeResidence*) ngoài nước Ấn độ hay không? Trả lời 'có' hoặc 'không' bằng 2 cách: dùng lệnh CASE hoặc hàm IF

- (12).- Cho biết tên quốc gia của những nhân viên trong table DS_Salaries có quốc gia cư trú (*EmployeeResidence*) trùng với quốc gia nơi đặt văn phòng chính của Công ty mà họ đang làm việc (*CompanyLocation*).

name
Algeria
American Samoa
Argentina
Armenia
...
(tổng cộng gồm 67 records)

- (13).- Cho biết số lượng người làm việc của từng mức làm việc online (*RemoteRatio*).

RemoteRatio	Quantity
0	1923
50	189
100	1643

- (14).- Cho biết lương bình quân tính theo USD của JobTitleName là 'AI Developer'.
- (15).- Cho biết trong năm 2021, những công ty có CompanySizeName là 'large' đã không trả lương cho những JobTitleName nào?
- (16).- Thống kê cho biết lương cao nhất theo từng năm của những người được trả bằng đơn vị tiền tệ là 'United Kingdom Pound'.
- (17).- Dựa trên quốc gia nơi đặt văn phòng chính của Công ty (*CompanyLocation*), hãy thực hiện thống kê dữ liệu trong table DS_Salaries theo vùng (field *region* trong table *CountriesWithRegionalCodes*) gồm lương bình quân theo dollar (*SalaryInUsd*), số lượng nhân viên thuộc vùng đó, Số lượng nhân viên có mức lương (*SalaryInUsd*) trên 100.000, số lượng công ty có CompanySize là 'Small'.

region	Avg Salary	SLNV	SLNV >100K	SL SmallCompany
Africa	53306.4	14	1	7
Americas	150178.0	3168	2563	64
Asia	45681.8	97	7	19
Europe	74895.2	458	90	55
Oceania	74084.5	18	4	3

- (18).- Tương tự câu trên, nhưng dựa trên quốc gia nơi cư trú của nhân viên (*EmployeeResidence*), hãy thực hiện thống kê dữ liệu trong table DS_Salaries theo vùng (field *region* trong table *CountriesWithRegionalCodes*) gồm lương bình quân theo dollar (*SalaryInUsd*), số lượng nhân viên thuộc vùng đó, Số lượng nhân viên có mức lương (*SalaryInUsd*) trên 100.000, số lượng công ty có CompanySize là 'Small'.

region	Avg Salary	SLNV	SLNV >100K	SL SmallCompany
Africa	49515.2	17	1	7
Americas	150915.9	3140	2559	55
Asia	50661.4	120	12	26
Europe	74852.4	464	90	58
Oceania	74881.5	14	3	2

- (19).- Cho biết tên chức danh (*JobTitleName*) có mức lương cao nhất theo từng năm.

work_year	JobTitleName	SalaryInUsd
2020	Research Scientist	450000
2021	Applied Machine Learning Scientist	423000
2022	Data Analyst	430967
2023	AI Scientist	423834

- (20).- Cho biết có phải chức danh ‘*Computer Vision Engineer*’ là chức danh có mức lương tính theo USD (*SalaryInUSD*) là cao nhất hay không? Chỉ trả lời có hoặc không.
- (21).- Cho biết lương bình quân tính theo USD (*SalaryInUsd*) của chức danh nghề nghiệp (*JobTitleName*) là ‘*Data Manager*’ tại Châu Phi (*Africa*) là bao nhiêu?
- (22).- Lần lượt sử dụng các hàm sau đây cho biết tên nước trả lương (*SalaryInUsd*) “cao nhất” hoặc “thấp nhất”:
- ANY_VALUE
 - MAX_BY/MIN_BY
 - Không sử dụng các hàm ANY_VALUE và MAX_BY/MIN_BY.

Kết quả truy vấn:

Có sử dụng các hàm ANY_VALUE và MAX_BY/MIN_BY		Không sử dụng các hàm ANY_VALUE và MAX_BY/MIN_BY	
max age	min age	max_salary	min_salary
NZL	TKM	United States of America	null
		null	Czechia

- (23).- Cho biết mức lương cao nhất (tính theo *SalaryInUSD*) của từng chức danh (*JobTitleName*) của những công ty có văn phòng chính đóng tại ‘*United Kingdom of Great Britain and Northern Ireland*’. Sắp xếp giảm dần theo MaxSalary. Minh họa kết quả cần thực hiện.

JobTitleName	MaxSalary
Data Analyst	430967
Analytics Engineer	230000
Marketing Data Analyst	200000
Research Scientist	195895
Machine Learning Engineer	195652
...	
(tổng cộng gồm 25 records)	

- (24).- Tạo thống kê sau:

region	LargeSize	MediumSize	SmallSize
Africa	3	7	7
Americas	280	2805	55
Asia	62	32	26
Europe	103	303	58
Oceania	6	6	2

- (25).- Giả sử có quy ước về *SalaryInUsd* như sau:

- Thuộc loại HighSalary khi *SalaryInUsd* > 200.000
- Thuộc loại MiddleSalary khi 150.000 <= *SalaryInUsd* <= 200.000
- Thuộc loại LowSalary khi *SalaryInUsd* < 150.000

Thống kê số lượng người thuộc mức lương tương ứng của từng loại và sắp xếp theo tên nước tăng dần như minh họa sau:

name	HighSalary	MiddleSalary	LowSalary
Albania	0	0	1
Algeria	0	1	0
American Samoa	0	1	2
Argentina	0	1	2
Armenia	0	1	0

name	HighSalary	MiddleSalary	LowSalary
...
(tổng cộng gồm 72 records)			

(26).- Tạo thống kê sau cho biết tên nước có số lượng người đạt mức cao nhất của từng loại mức lương:

name	HighSalary	MiddleSalary	LowSalary
United States of America	508	2502	0
India	0	0	50
...

(27).- Cho biết tên các Châu lục (Region) và số lượng các tiểu vùng của châu lục (sub-region) được ghi nhận trong table *CountriesWithRegionalCodes*.

region	NumberOfSub_Region
null	1
Africa	2
Americas	2
Asia	5
Europe	4
Oceania	4

(28).- Cho biết lương bình quân của từng loại EmploymentType được trả tại Canada là bao nhiêu? Yêu cầu: hiện đầy đủ EmploymentTypeName của tất cả các EmploymentType.

Minh họa kết quả cần thực hiện

EmploymentTypeName	MaxSalary
Full-time	\$275,000.00
Consultants	do not use this employment type
Freelancer	\$100,000.00
Part-time employees	do not use this employment type

(29).- Cho biết tên chức danh (JobTitleName) có mức lương cao nhất của những công ty có văn phòng chính đóng tại 'United Kingdom of Great Britain and Northern Ireland'. Minh họa kết quả cần thực hiện, yêu cầu thực hiện bằng 2 cách:

- Sử dụng mệnh đề LIMIT
- Không sử dụng mệnh đề LIMIT

JobTitleName	MaxSalary
Data Analyst	430967

(30).- Thống kê cho các quốc gia thuộc khu vực Châu Á (region='Asia') với thông tin gồm: SalaryLevel (làm tròn đến giá trị ngàn USD), chức danh nghề nghiệp (JobTitleName), hình thức làm việc (EmploymentTypeName), và số lượng người trong từng nhóm. Sắp xếp kết quả tăng dần lần lượt theo 3 cột đầu của kết quả.

Salary Level	JobTitleName	EmploymentTypeName	NumberOf Employees
5000.0	3D Computer Vision Researcher	Part-time employees	1
5000.0	Power BI Developer	Full-time	1
6000.0	Data Analyst	Full-time	3
6000.0	Data Science Consultant	Full-time	1
6000.0	Product Data Analyst	Full-time	1
8000.0	BI Data Analyst	Full-time	1
8000.0	Data Analyst	Full-time	1
8000.0	Product Data Scientist	Full-time	1
...

BÀI TẬP SỐ 4

(Common Table Expression - CTE)

4.1. Recursive Common table expression (Recursive CTE)

1. Sử dụng recursive CTE để tính các tổng / tích sau. Tính tổng các số từ 1-n (giả sử với n=10) cho các trường hợp sau:

n	tong	n	tong	n	tong	n	tong
1	1	0	1	1	1.0	0	0.0
2	3	1	4	2	1.5	1	0.5
3	6	2	9	3	1.833	2	0.833
4	10	3	16	4	2.083	3	1.083
5	15	4	25	5	2.283	4	1.283
6	21	5	36	6	2.45	5	1.45
7	28	6	49	7	2.593	6	1.593
8	36	7	64	8	2.718	7	1.718
9	45	8	81	9	2.829	8	1.829
10	55	9	100	10	2.829	9	1.929
		10	121		2.929	10	2.02
<i>a</i>		<i>b</i>		<i>c</i>		<i>d</i>	
Minh họa kết quả thực hiện của các câu từ 1.a đến 1.d							

a.- $S = 1 + 2 + 3 + \dots + n$

b.- $S = 1 + 3 + 5 + \dots + (2n+1)$

c.-

$$S = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}$$

d.-

$$S = \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots + \frac{1}{n+1}$$

e.-

$$S = \frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \dots + \frac{1}{2n}$$

f.-

$$S = 1 + \frac{1}{3} + \frac{1}{5} + \dots + \frac{1}{2n-1}$$

g.-

$$S = \frac{1}{2} + \frac{2}{3} + \frac{3}{4} + \dots + \frac{n}{n+1}$$

h.-

$$S = \frac{1}{2} + \frac{3}{4} + \frac{5}{6} + \dots + \frac{2n+1}{2n+2}$$

i.- $P = 1 \times 2 \times 3 \times \dots \times n$

j.- $P = 1 \times 3 \times 5 \times \dots \times (2n-1)$

2. Tìm chữ số lớn nhất trong 1 số nguyên dương n, với giá trị của n là tùy ý. VD: n=29730 => in ra số 9 theo hình minh họa:
3. Tính tổng các chữ số của số nguyên dương n. VD: n=29730=> 21
4. Tính tổng các chữ số chẵn của số nguyên dương n.
5. Tính tích các chữ số lẻ của số nguyên dương n.
6. Cho số nguyên dương n. Tìm chữ số đầu tiên của n.
7. Đếm số lượng chữ số lẻ của số nguyên dương n.
8. Dãy Fibonacci {Fn} bậc 2 được định nghĩa: $f_0 = f_1 = 1$
 $f_n = f_{n-1} + f_{n-2} \quad (n \geq 2)$

SoLonNhat	▼
9	

Sử dụng CTE để tạo kết quả chứa các giá trị trong dãy Fibonacci như sau:

a.- Dãy f1 có giá trị từ 0 đến 1.000.

b.- Dãy f1 gồm 11 số.

n	DemSoLe
2911730	0
291173	0
29117	1
2911	2
291	3
29	4
2	5
0	5

Bài 7

	f1	f2
1	0	1
2	1	1
3	1	2
4	2	3
5	3	5
6	5	8
7	8	13
8	13	21
9	21	34
10	34	55
11	55	89
12	89	144
13	144	233
14	233	377
15	377	610
16	610	987
17	987	1597

Bài 8a

	sl	f1	f2
1	0	0	1
2	1	1	1
3	2	1	2
4	3	2	3
5	4	3	5
6	5	5	8
7	6	8	13
8	7	13	21
9	8	21	34
10	9	34	55
11	10	55	89

Bài 8b

9. Cho table Employees cho biết thông tin về nhân sự trong 1 công ty như sau:

- Dữ liệu ban đầu:

EmployeeID	FirstName	LastName	BossID
1	Domenic	Leaver	5
2	Cleveland	Hewins	1
3	Kakalina	Atherton	8
4	Roxanna	Fairlie	
5	Hermie	Comsty	4
6	Pooh	Goss	8
7	Faulkner	Challiss	5
8	Bobbe	Blakeway	4
9	Laurene	Burchill	1
10	Augusta	Gosdin	8

- Minh họa kết quả cần thực hiện:

EmployeeID	Employee_Name	Position	BossID	Boss_Name
4	Roxanna Fairlie	Manager		
8	Bobbe Blakeway	Project Leader	4	Roxanna Fairlie
5	Hermie Comsty	Project Leader	4	Roxanna Fairlie
1	Domenic Leaver	Team Leader	5	Hermie Comsty
7	Faulkner Challiss	Team Leader	5	Hermie Comsty
6	Pooh Goss	Team Leader	8	Bobbe Blakeway
3	Kakalina Atherton	Team Leader	8	Bobbe Blakeway
10	Augusta Gosdin	Team Leader	8	Bobbe Blakeway
2	Cleveland Hewins	Programmer	1	Domenic Leaver

Thuộc tính BossID trong table cho biết mã số của sếp (boss) trực tiếp của nhân viên (EmployeeID) tương ứng. Do thuộc tính BossID của Roxana là NULL, cho biết Roxana là chủ tịch hoặc chủ sở hữu của công ty.

Hãy sử dụng CTE để quy để liệt kê tên + họ, chức vụ của tất cả nhân viên và tên+họ sếp trực tiếp của nhân viên đó.

Biết rằng trong công ty chỉ bao gồm các chức vụ (position) từ cao xuống thấp như sau: Manager, Project Leader, Team Leader, Programmer.

10. Cho table CTE_Data cho biết khoảng cách giữa các thành phố của nước Đức như sau:

- Dữ liệu ban đầu:

city_from	city_to	distance
Groningen	Heerenveen	61.4
Groningen	Harlingen	91.6
Harlingen	Wieringerwerf	52.3
Wieringerwerf	Hoorn	26.5
Hoorn	Amsterdam	46.1
Amsterdam	Haarlem	30
Heerenveen	Lelystad	74
Lelystad	Amsterdam	57.2

- Yêu cầu sử dụng CTE để tìm tất cả những khả năng từ thành phố 'Groningen' đến thành phố 'Haarlem' cùng cự ly cần di chuyển của mỗi phương án.

- Minh họa kết quả cần thực hiện:

route ▼	RouteLength ▼
Groningen -> Heerenveen -> Lelystad -> Amsterdam -> Haarlem	222.6
Groningen -> Harlingen -> Wieringerwerf -> Hoorn -> Amsterdam -> Haarlem	246.5

4.2. Athletes

Yêu cầu: Các câu trong phần này, nếu được, ngoài cách sử dụng CTE, học viên có thể viết thêm cách khác mà không sử dụng CTE.

11. Gọi X là danh sách các nước có VDV đạt huy chương bạc đã được sắp xếp tăng hoặc giảm dần) theo số lượng VDV đạt huy chương bạc. Liệt kê các nước có huy chương bạc nhưng không nằm trong 10 nước có số lượng VDV đạt huy chương bạc nhiều nhất và cũng không nằm trong 10 nước có số lượng VDV đạt huy chương bạc ít nhất.
12. Cho biết số lượng VDV nữ của nước Qatar (QAT)) có đông hơn số lượng VDV nữ của nước Ghana (GHA) hay không? Trả lời có hoặc không.
13. Cho biết tên những môn thi (sport) mà Việt Nam không tham gia.
14. Gọi R là tỷ lệ VDV có giải/tổng số VDV tham dự của mỗi nước trong môn fencing (vượt rào). Cho biết 3 nước có tỷ lệ R là cao nhất.
15. Cho biết nước có tỷ lệ VDV nữ/VDV nam là thấp nhất.
16. Thống kê số lượng VDV nam, nữ theo từng môn thi đấu (sport), sau đó chọn ra 5 môn có số lượng nữ và 5 môn có số lượng nam tham gia là nhiều nhất.

Minh họa kết quả cần thực hiện

nationality ▼	ratio winning_participating ▼	nationality ▼	Ratio_Female_Male ▼	sport ▼	SLNu ▼	SLNam ▼
ROU	66.67	QAT	5.41	athletics	1137	1226
RUS	60.0	HON	7.14	aquatics	716	729
HUN	60.0	ERI	9.09	football	264	347
USA	52.94			rowing	216	331
FRA	42.11			hockey	216	216
ITA	41.18					

Câu 14

Câu 15

Câu 16

17. (*)Cho biết những nước tham gia thi đấu tất cả những môn mà đoàn nước Maroc (MAR – moroccan) đã tham gia

4.3. Data Science salary

18. Cho biết số lượng của từng EmploymentType có trong mỗi loại CompanySize. Kết quả như minh họa sau:

SizeName ▼	NumOfFT ▼	NumOfCT ▼	NumOfFL ▼	NumOfPT ▼
Small	134	4	4	6
Large	445	2	1	6
Medium	3139	4	5	5

- 19.** Trong các EmploymentType ở Ấn Độ (EmployeeResidence= 'IN'), cho biết EmploymentType nào có số lượng là nhiều nhất và EmploymentType nào có số lượng là ít nhất?

(a).- Trường hợp chỉ lấy 1 dòng có số lượng nhiều (hoặc ít) nhất làm đại diện.

max_employmenttype	min_employmenttype
FT	PT

(b).- Trường hợp lấy tất cả các dòng cùng có số lượng nhiều (hoặc ít) nhất

max_type	min_type
Full-time	null
null	Part-time employees
null	Freelancer
null	Consultants

- 20.** Trong các JobTitleName ở Anh (EmployeeResidence= 'GB'), cho biết JobTitleName có lương bình quân là thấp nhất?
- 21.** Thống kê tỷ lệ NV làm việc online (dựa trên field Remote Ratio) của từng mức kinh nghiệm của nhân viên (*Experience*).
- 22.** Nếu xem EmployeeResidence là nơi cư trú của người lao động. Nơi cư trú nào mà mức chênh lệch tiền lương giữa người có mức lương cao nhất và người có lương thấp nhất (dựa trên SalaryInUSD) là lớn nhất.
- 23.** Cho biết lương của JobTitleName là "Data Analyst" ở Ấn Độ (Employee='IN') có cao hơn cùng chức danh đó nhưng ở Anh ('GB') hay không?
- 24.** Cho biết tên nước nào có chênh lệch giữa lương cao nhất và lương thấp nhất là lớn nhất.
- 25.** Cho biết tỷ lệ của từng loại Experience so với tổng số khảo sát chỉ tính cho những công ty có văn phòng đóng tại Mỹ (Company Location='US').

BÀI TẬP SỐ 5 (Window Functions)

5.1. Athletes

5.1.1. Numbering functions

(1)- Xếp hạng (ranking) các môn thi (sport) theo số lượng VDV tham gia với yêu cầu:

- Các môn thi đấu có cùng số lượng VDV tham gia sẽ có cùng hạng.
- Khi có các đội có cùng hạng thì đội ngay sau đó không được lấy hạng liền kề (trong kết quả sau 2 môn table tennis và badminton có cùng hạng 22 nên môn taekwondo phải nhận hạng 24. Tương tự, 2 môn taekwondo và archery có cùng hạng 24 nên môn golf phải lấy hạng 26).

SPORT	NumberOfParticipatingAthletes	Ranking
athletics	2363	1
aquatics	1445	2
...
tennis	196	21
table tennis	172	22
badminton	172	22
taekwondo	128	24
archery	128	24
golf	120	26
...

(2)- Xếp hạng (ranking) các quốc gia dựa trên tổng số huy chương vàng, bạc đồng mà các VDV của quốc gia đó đã đạt được. Lưu ý việc xếp hạng dựa trên: 3 tiêu chí lần lượt là tổng số huy chương vàng, tổng số huy chương bạc, tổng số huy chương đồng sao cho nếu số huy chương của cả 3 loại bằng nhau thì thứ hạng phải bằng nhau (như trong minh họa 2 nước VIE và BRN)

nationality	Gold	Silver	Bronze	Ranking
USA	139	54	71	1
GBR	64	55	26	2
RUS	52	29	34	3
...
CZE	1	2	12	46
ETH	1	2	5	47
SLO	1	2	1	48
VIE	1	1	0	49
BRN	1	1	0	49
...

(3)- Xếp hạng các VDV dựa trên môn thi đấu (sport) và tổng số huy chương vàng, bạc đồng mà các VDV đã đạt được. Lưu ý việc xếp hạng:

- Dựa trên 3 tiêu chí lần lượt là tổng số huy chương vàng, tổng số huy chương bạc, tổng số huy chương đồng sao cho nếu số huy chương của cả 3 loại bằng nhau thì thứ hạng phải bằng nhau.
- Nếu có những VDV đồng hạng thì thứ hạng của VDV sau đó sẽ là: thứ hạng của VDV liền trước + số lượng VDV có cùng hạng đó +1.
- Thứ hạng của mỗi môn thi đấu sẽ bắt đầu tính từ 1 trở đi.

- Minh họa kết quả cần thực hiện:

sport	name	Gold	Silver	Bronze	Ranking
aquatics	Michael Phelps	5	1	0	1
aquatics	Katie Ledecky	4	1	0	2
...
aquatics	Nathan Adrian	2	0	2	7
aquatics	Tingmao Shi	2	0	0	8
aquatics	Natalia Ishchenko	2	0	0	8
aquatics	Aisen Chen	2	0	0	8
aquatics	Anthony Ervin	2	0	0	8
aquatics	Svetlana Romashina	2	0	0	8
aquatics	Lilly King	2	0	0	8
aquatics	Caeleb Dressel	2	0	0	8
aquatics	Emma McKeon	1	2	1	15
...
aquatics	Henrik Christiansen	0	0	0	21
archery	Hyejin Chang	2	0	0	1
archery	Bonchan Ku	2	0	0	1
archery	Bobae Ki	1	0	1	3
archery	Seungyun Lee	1	0	0	3
archery	Misun Choi	1	0	0	3
...

(4)- Xếp hạng các VDV theo mỗi nước dựa trên tổng số huy chương vàng, bạc đồng mà các VDV đã đạt được. Lưu ý việc xếp hạng:

- Nếu số huy chương vàng của 2 VDV bằng nhau thì xét tiếp trên số số huy chương bạc; nếu cả huy chương vàng và bạc cùng bằng nhau thì xét tiếp trên số huy chương đồng.
- Nếu những VDV có cùng số lượng huy chương mỗi loại như nhau thì phải được xếp đồng hạng.
- Thứ hạng của mỗi môn thi đấu sẽ bắt đầu tính từ 1 trở đi.
- Chỉ hiện danh sách các VDV có thứ hạng từ 1 đến 3.
- Nếu 1 nước có nhiều VDV có cùng hạng 1(hoặc 2 hay 3) sẽ dẫn đến số lượng VDV của nước đó vượt quá 3 người. Học viên cần loại bỏ những người thừa này.
- Minh họa kết quả mà học viên có thể gặp (chưa đúng với yêu cầu do số lượng VDV của 1 nước có thể vượt quá 3 người như 2 nước UZB và TUR trong hình minh họa sau):

nationality	Ranking	name	sport	Gold	Silver	Bronze
VIE	1	Xuan Vinh Hoang	shooting	1	1	0
VEN	1	Yulimar Rojas	athletics	0	1	0
VEN	2	Yoel Segundo Finol	boxing	0	0	1
VEN	2	Stefany Hernandez	cycling	0	0	1
UZB	1	Shakhobidin Zoirov	boxing	1	0	0
UZB	1	Fazliddin Gaibnazarov	boxing	1	0	0
UZB	1	Hasanboy Dusmatov	boxing	1	0	0
UZB	1	Ruslan Nurudinov	weightlifting	1	0	0
USA	1	Michael Phelps	aquatics	5	1	0
USA	2	Katie Ledecky	aquatics	4	1	0
USA	3	Simone Biles	gymnastics	4	0	1
UKR	1	Oleg Verniaiev	gymnastics	1	1	0

UKR	2	Iurii Cheban	canoe	1	0	0
UKR	3	Olga Kharlan	fencing	0	1	1
UAE	1	Sergiu Toma	judo	0	0	1
TUR	1	Taha Akgul	wrestling	1	0	0
TUR	2	Riza Kayaalp	wrestling	0	1	0
TUR	2	Selim Yasar	wrestling	0	1	0
TUR	2	Daniyar Ismayilov	weightlifting	0	1	0

- Minh họa kết quả đúng mà học viên cần thực hiện:

nationality	Ranking	name	sport	Gold	Silver	Bronze
VIE	1	Xuan Vinh Hoang	shooting	1	1	0
VEN	1	Yulimar Rojas	athletics	0	1	0
VEN	2	Yoel Segundo Finol	boxing	0	0	1
VEN	2	Stefany Hernandez	cycling	0	0	1
UZB	1	Shakhobidin Zoirov	boxing	1	0	0
UZB	1	Fazliddin Gaibnazarov	boxing	1	0	0
UZB	1	Hasanboy Dusmatov	boxing	1	0	0
USA	1	Michael Phelps	aquatics	5	1	0
USA	2	Katie Ledecky	aquatics	4	1	0
USA	3	Simone Biles	gymnastics	4	0	1
UKR	1	Oleg Verniaiev	gymnastics	1	1	0
UKR	2	Iurii Cheban	canoe	1	0	0
UKR	3	Olga Kharlan	fencing	0	1	1
UAE	1	Sergiu Toma	judo	0	0	1
TUR	1	Taha Akgul	wrestling	1	0	0
TUR	2	Riza Kayaalp	wrestling	0	1	0
TUR	2	Selim Yasar	wrestling	0	1	0
...

5.1.2. Navigation functions

(5)- Thống kê danh sách theo mẫu, với các yêu cầu:

- Kết quả chỉ bao gồm tên những VDV có huy chương.
- Danh sách được tổ chức theo từng quốc gia, trong mỗi quốc gia sẽ dựa trên 3 tiêu chí lần lượt là tổng số huy chương vàng, tổng số huy chương bạc, tổng số huy chương đồng.
- Cột cuối cùng sẽ lấy tên VDV có thành tích cao nhất của nước đó.

➤ **Yêu cầu:** lần lượt thực hiện 3 cách khác nhau, mỗi cách sử dụng 1 trong 3 hàm sau: *FIRST_VALUE*, *LAST_VALUE*, *nTh_VALUE*

nationality	name	Gold	Silver	Bronze	Ranking
VIE	Xuan Vinh Hoang	1	1	0	Xuan Vinh Hoang
VEN	Yulimar Rojas	0	1	0	Yulimar Rojas
VEN	Yoel Segundo Finol	0	0	1	Yulimar Rojas
VEN	Stefany Hernandez	0	0	1	Yulimar Rojas
UZB	Shakhobidin Zoirov	1	0	0	Shakhobidin Zoirov
UZB	Fazliddin Gaibnazarov	1	0	0	Shakhobidin Zoirov
UZB	Hasanboy Dusmatov	1	0	0	Shakhobidin Zoirov
UZB	Ruslan Nurudinov	1	0	0	Shakhobidin Zoirov
UZB	Shakhram Giyasov	0	1	0	Shakhobidin Zoirov

nationality	name	Gold	Silver	Bronze	Ranking
UZB	Bektemir Melikuziev	0	1	0	Shakhobidin Zoirov
UZB	Diyorbek Urozboev	0	0	1	Shakhobidin Zoirov
UZB	Rishod Sobirov	0	0	1	Shakhobidin Zoirov
UZB	Rustam Tulaganov	0	0	1	Shakhobidin Zoirov
UZB	Murodjon Akhmadaliev	0	0	1	Shakhobidin Zoirov
UZB	Ikhtiyor Navruzov	0	0	1	Shakhobidin Zoirov
UZB	Magomed Idrisovitch Ibragimov	0	0	1	Shakhobidin Zoirov
UZB	Elmurat Tasmuradov	0	0	1	Shakhobidin Zoirov
USA	Michael Phelps	5	1	0	Michael Phelps
USA	Katie Ledecky	4	1	0	Michael Phelps
...

5.1.3. Analytics functions

(6)- Viết truy vấn để có kết quả như minh họa sau bằng 2 cách: có và không có sử dụng window function, trong đó:

- `athletes_GoldMedal` : số lượng VDV đạt huy chương vàng của nước đó.
- `athletes_SilverMedal`: số lượng VDV đạt huy chương bạc của nước đó.
- `athletes_BronzeMedal`: số lượng VDV đạt huy chương đồng của nước đó.

Nhận xét: khi 1 VDV vừa đạt huy chương vàng, vừa đạt huy chương bạc thì VDV này sẽ được đếm 2 lần: 1 lần đếm cho huy chương vàng và 1 lần đếm cho huy chương bạc. Tương tự đối với những trường hợp khác.

nationality	athletes_GoldMedal	athletes_SilverMedal	athletes_BronzeMedal
USA	119	50	70
GBR	59	51	26
RUS	49	26	34
GER	46	42	67
CHN	42	28	37
...

(7)- Viết truy vấn để có kết quả như minh họa sau

↪ **Giải thích:**

- `FemaleGold` : số lượng VDV nữ đạt huy chương vàng của từng môn thi đấu.
- `Female` : tổng số lượng VDV nữ tham gia thi đấu đó.
- `FemaleRatio` : Tỷ lệ VDV nữ đạt huy chương vàng trên tổng số VDV nữ tham gia môn thi đấu đó (= `FemaleGold` / `Female`)

↪ **Yêu cầu:**

- Lần lượt thực hiện bằng 2 cách: có và không có sử dụng window function.
- Cần thể hiện dấu phần trăm (%) cho cột `FemaleRatio`.
- Danh sách được sắp xếp giảm dần theo giá trị số của cột `FemaleRatio`.

Row	sport	FemaleGold	Female	FemaleRatio
1	rowing	20	216	9.26%
2	basketball	12	144	8.33%
3	boxing	3	36	8.33%
4	handball	15	180	8.33%
5	rugby sevens	12	148	8.11%
6	fencing	10	124	8.06%
7	aquatics	54	716	7.54%

(8)- Viết truy vấn để có kết quả như minh họa sau

➤ **Giải thích:**

- FemaleGold : số lượng VDV nữ đạt huy chương vàng của nước đó.
- Female : tổng số lượng VDV nữ của nước đó.
- FemaleRatio : Tỷ lệ VDV nữ đạt huy chương vàng trên tổng số VDV nữ của nước đó (= FemaleGold / Female).
- MaleGold : số lượng VDV nam đạt huy chương vàng của nước đó.
- Male : tổng số lượng VDV nam của nước đó.
- MaleRatio : Tỷ lệ VDV nam đạt huy chương vàng trên tổng số VDV nam của nước đó (= MaleGold / Male)

➤ **Yêu cầu:**

- Lần lượt thực hiện bằng 2 cách: có và không có sử dụng window function.
- Cần thể hiện dấu phần trăm (%) cho 2 cột FemaleRatio và MaleRatio.
- Danh sách được sắp xếp giảm dần theo giá trị số của cột FemaleRatio. Nếu bằng nhau sẽ sắp xếp dựa trên giá trị số của cột MaleRatio.

<i>nationality</i>	<i>FemaleGold</i>	<i>Female</i>	<i>FemaleRatio</i>	<i>MaleGold</i>	<i>Male</i>	<i>MaleRatio</i>
RUS	38	142	26.76%	11	144	7.64%
USA	74	303	24.42%	45	264	17.05%
KOS	1	5	20%	0	3	0%
GBR	27	169	15.98%	32	205	15.61%
GER	29	204	14.22%	17	237	7.17%
CHN	28	251	11.16%	14	153	9.15%

(9)- Viết truy vấn để có kết quả như minh họa sau

➤ **Giải thích:**

- AthletesGoldMedal : số lượng VDV đạt huy chương vàng của nước đó. Kết quả truy vấn sẽ được sắp xếp giảm dần trên cột này.
- AVG_3nationality : là giá trị trung bình của tổng số huy chương vàng của nước đang xét và 1 nước ngay trước cùng 1 nước ngay sau nước đó.

<i>nationality</i>	<i>athletes_GoldMedal</i>	<i>AVG_3_nationality</i>
USA	119	89.0
GBR	59	75.67
RUS	49	51.33
GER	46	45.67
CHN	42	41.67
BRA	37	34.0

(10)- Viết truy vấn để có kết quả như minh họa sau:

➤ **Giải thích:**

- GoldMedal : số lượng huy chương vàng mà nước đó đạt được trong môn (sport) tương ứng.
- nationality_max_gold : là thông tin về nước có số lượng huy chương vàng cao nhất của môn đó, gồm tên nước và số lượng huy chương vàng mà nước đó đạt được.

<i>sport</i>	<i>nationality</i>	<i>GoldMedal</i>	<i>nationality_max_gold</i>
aquatics	USA	61	USA - 61 Gold medal
aquatics	SRB	13	USA - 61 Gold medal
aquatics	RUS	11	USA - 61 Gold medal

sport	nationality	GoldMedal	nationality_max_gold
...
aquatics	SIN	1	USA - 61 Gold medal
archery	KOR	8	KOR - 8 Gold medal
athletics	USA	27	USA - 27 Gold medal
athletics	JAM	11	USA - 27 Gold medal
athletics	KEN	6	USA - 27 Gold medal
...

5.2. Data Science Salary

5.2.1. Numbering functions

(11)- Xếp hạng theo mức lương cao nhất (dựa trên thuộc tính *SalaryInUsd*) của từng quốc gia (*CompanyLocation*)

<i>name</i>	<i>MaxSalaryInUsd</i>	<i>RankingByMaxSalary</i>
United States of America	450000	1
United Kingdom of Great Britain and Northern Ireland	430967	2
Israel	423834	3
India	300000	4
Canada	275000	5
Germany	275000	5
Japan	260000	7
...

(12)- Xếp hạng theo mức lương trung bình (dựa trên thuộc tính *SalaryInUsd*) theo từng chức danh công việc (*JobTitleName*)

<i>JobTitleName</i>	<i>AvgSalaryInUsd</i>	<i>RankingByAvgSalary</i>
Data Lead	USD 212,500.00	3
Data Analytics Lead	USD 211,254.50	4
Principal Data Scientist	USD 198,171.13	5
Director of Data Science	USD 195,140.73	6
Principal Data Engineer	USD 192,500.00	7

Yêu cầu: cột *AvgSalaryInUsd* được định dạng có 2 số lẻ cùng dấu ngăn cách phần ngàn và có từ 'USD' đi liền trước số tiền.

5.2.2. Navigation functions

(13)- Lập truy vấn để có kết quả như minh họa sau. Lưu ý định dạng dữ liệu như minh họa:

<i>JobTitleName</i>	<i>SalaryInUsd</i>	<i>Highest salary of the same title</i>
3D Computer Vision Researcher	\$50,000.00	\$50,000.00
3D Computer Vision Researcher	\$20,000.00	\$50,000.00
3D Computer Vision Researcher	\$10,000.00	\$50,000.00
3D Computer Vision Researcher	\$5,409.00	\$50,000.00
AI Developer	\$300,000.00	\$300,000.00
AI Developer	\$275,000.00	\$300,000.00
...

➤ Yêu cầu: lần lượt thực hiện 3 cách khác nhau, mỗi cách sử dụng 1 trong 3 hàm sau: *FIRST_VALUE*, *LAST_VALUE*, *nTh_VALUE*

5.2.3. Analytics functions

(14)- Thống kê lương (*SalaryInUsd*) theo từng chức danh nghề nghiệp, với yêu cầu:

- *Average salary of the same title*: là lương trung bình của những người có cùng *JobTitleName*.
- *Salary difference* = *Average salary of the same title* - *SalaryInUsd*
- Kết quả được sắp xếp tăng dần theo *JobTitleName*.
- Định dạng cột *SalaryInUsd* có 1 số lẻ.
- Định dạng cột *Average salary of the same title* có 2 số lẻ và có đơn vị tiền tệ là '\$'.

<i>JobTitleName</i>	<i>SalaryInUsd</i>	<i>Average salary of the same title</i>	<i>Salary difference</i>
3D Computer Vision Researcher	5,409.0	\$21,352.25	15943.25
3D Computer Vision Researcher	10,000.0	\$21,352.25	11352.25
3D Computer Vision Researcher	20,000.0	\$21,352.25	1352.25
3D Computer Vision Researcher	50,000.0	\$21,352.25	-28647.75
AI Developer	6,304.0	\$136,666.09	130362.09
AI Developer	60,000.0	\$136,666.09	76666.09
...

(15)- Thống kê lương (*SalaryInUsd*) theo từng quốc gia của người lao động (*EmployeeResidence*) qua từng năm (*work_year*), với yêu cầu:

- *MaxSalaryOfYear*: là lương cao nhất của năm tương ứng.
- *HighestSalaryInYears*: trong những năm của từng quốc gia, mức lương cao nhất là bao nhiêu?
- *YearHasMaxSalary*: mức lương cao nhất (*HighestSalaryInYears*) là vào năm nào?

<i>Country of EmployeeResidence</i>	<i>work_year</i>	<i>MaxSalary OfYear</i>	<i>Highest SalaryInYears</i>	<i>YearHas MaxSalary</i>
Algeria	2022	100000	100000	2022
American Samoa	2021	45555	45555	2021
Argentina	2022	60000	60000	2022
Argentina	2023	30000	60000	2022
Armenia	2023	50000	50000	2023
Australia	2021	75050	150000	2022
Australia	2022	150000	150000	2022
Australia	2023	68318	150000	2022
Austria	2020	91237	91237	2020
Austria	2022	89306	91237	2020
Austria	2023	60093	91237	2020
...

BÀI TẬP SỐ 6 (FROM - Pivot & UnPivot)

Yêu cầu đối với phần Pivot

- Học viên hãy lần lượt thực hiện mỗi yêu cầu bằng 2 cách: có và không có dùng FROM ... PIVOT.
- Đối với các truy vấn có sử dụng Pivot, học viên lưu kết quả thực hiện của truy vấn thành 1 table mới theo quy ước: câu 1 lưu kết quả vào table tên 'one', câu 2 lưu kết quả vào table 'two', ...

6.1. Athletes

6.1.1. Pivot

- (1)- Tạo thống kê số lượng VDV nam và nữ của từng quốc gia, sắp xếp kết quả giảm dần theo số lượng VDV nữ.

<i>nationality</i>	<i>male</i>	<i>female</i>
USA	264	303
CHN	153	251
AUS	211	220
BRA	269	216
GER	237	204
...

- (2)- Tạo thống kê số lượng VDV theo từng năm sinh (từ năm 1980 đến năm 1990), sắp xếp kết quả giảm dần theo quốc gia.

<i>nationality</i>	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
ZIM	1	0	1	1	0	1	3	3	3	1	4
ZAM	0	0	0	0	0	0	0	0	3	0	0
YEM	0	0	0	0	0	0	0	0	0	0	0
VIN	0	0	0	0	0	0	1	0	0	0	0
VIE	0	0	0	1	1	1	0	1	1	1	1
VEN	1	4	1	3	5	4	8	11	1	6	8
VAN	0	0	0	0	0	0	0	0	0	0	0
UZB	0	1	2	1	1	1	7	4	5	5	5
USA	6	13	12	11	15	35	35	26	38	58	51

- (3)- Tạo thống kê số lượng VDV theo các mức làm tròn của chiều cao gồm các mức: 200cm, 210cm, 220cm, 230cm. Với các yêu cầu:

- Làm tròn chiều cao theo mét với 1 số lẻ và chuyển sang đơn vị Cm.
- Làm tròn cân nặng đến hàng chục.
- Chỉ xét các mức chiều cao: 200cm, 210cm, 220cm, 230cm. Kết quả truy vấn chỉ chứa những nước có ít nhất 1 VDV thuộc một trong những chiều cao này.
- Sắp xếp kết quả giảm dần theo quốc gia.

Minh họa kết quả cần thực hiện:

<i>nationality</i>	<i>Height_200</i>	<i>Height_210</i>	<i>Height_220</i>	<i>Height_230</i>
ZAM	1	0	0	0
VEN	8	2	0	0

<i>nationality</i>	<i>Height_200</i>	<i>Height_210</i>	<i>Height_220</i>	<i>Height_230</i>
UZB	2	0	0	0
USA	45	5	1	0
UKR	6	1	0	0
TUR	1	0	0	0
TUN	5	0	0	0
TTO	2	0	0	0
...

(4)- Tạo thống kê số lượng VDV theo các môn thi đấu (sport) và theo trọng lượng của VDV tham gia thi đấu môn đó theo minh họa sau. Với các yêu cầu:

- Làm tròn chiều cao theo mét với 1 số lẻ và chuyển sang đơn vị Cm (*HeightInCm*).
- Làm tròn cân nặng đến hàng chục.
- Chỉ xét những môn có VDV tham gia có chiều cao từ 2 mét trở lên.
- Chỉ quan tâm các mức cân nặng từ 100 Kg – 140 Kg
- Kết quả truy vấn chỉ chứa những nước có ít nhất 1 VDV thuộc những chiều cao này và phải có số lượng VDV có cân nặng ở mức 120Kg >= 0.
- Sắp xếp kết quả tăng dần theo sport.

<i>sport</i>	<i>HeightInCm</i>	<i>Weight_100</i>	<i>Weight_110</i>	<i>Weight_120</i>	<i>Weight_130</i>	<i>Weight_140</i>
aquatics	200.0	29	16	5	1	0
athletics	200.0	5	6	10	6	7
athletics	210.0	0	0	3	0	0
basketball	200.0	28	15	2	0	0
basketball	210.0	9	13	16	2	1
basketball	220.0	1	0	1	0	0
canoe	200.0	2	0	1	0	0
handball	210.0	0	2	2	0	0
volleyball	210.0	11	3	1	0	0

(5)- Tạo thống kê số lượng VDV của môn thi đấu (sport) là 'cycling' theo trọng lượng của VDV tham gia thi đấu môn đó theo minh họa sau. Với các yêu cầu:

- Làm tròn chiều cao theo mét với 1 số lẻ và chuyển sang đơn vị Cm.
- Làm tròn cân nặng đến hàng chục.

<i>HeightInCm</i>	<i>Weight_60</i>	<i>Weight_70</i>	<i>Weight_80</i>	<i>Weight_90</i>	<i>Weight_100</i>
160	6	1	1	0	0
170	14	6	0	0	0
180	4	25	5	4	1
190	0	4	2	6	0
200	0	0	0	0	0

6.1.2. UnPivot

Yêu cầu: Tương ứng với mỗi câu trong phần Unpivot, HV cần thực hiện bằng 2 cách:

- *Cách 1:* sử dụng Unpivot từ table đã tạo trong phần Pivot
- *Cách 2:* không sử dụng Unpivot, viết lệnh truy vấn trực tiếp trên table athletes. Lúc này, thứ tự của các dòng (records) của **kết quả thực hiện có thể khác với khi dùng Unpivot.**

(6)- Từ table one của câu 1, viết lệnh truy vấn để có kết quả như minh họa sau:

Table one			Kết quả cần thực hiện		
nationality	male	female	nationality	quantity	sex
USA	264	303	USA	264	male
CHN	153	251	USA	303	female
AUS	211	220	CHN	153	male
BRA	269	216	CHN	251	female
GER	237	204	AUS	211	male
...	AUS	220	female
			BRA	269	male
			BRA	216	female
		

(7)- Từ table two của câu 2, viết lệnh truy vấn để có kết quả như minh họa sau:

• Table two

nationality	_1980	_1981	_1982	_1983	_1984	_1985	_1986	_1987	_1988	_1989	_1990
ZIM	1	0	1	1	0	1	3	3	3	1	4
ZAM	0	0	0	0	0	0	0	0	3	0	0
YEM	0	0	0	0	0	0	0	0	0	0	0
VIN	0	0	0	0	0	0	1	0	0	0	0
VIE	0	0	0	1	1	1	0	1	1	1	1
VEN	1	4	1	3	5	4	8	11	1	6	8
VAN	0	0	0	0	0	0	0	0	0	0	0
UZB	0	1	2	1	1	1	7	4	5	5	5
USA	6	13	12	11	15	35	35	26	38	58	51

• Kết quả cần thực hiện

nationality	Quantity	BirthYear
ZIM	1	1980
ZIM	0	1981
ZIM	1	1982
ZIM	1	1983
ZIM	0	1984
ZIM	1	1985
ZIM	3	1986
ZIM	3	1987
ZIM	3	1988
ZIM	1	1989
ZIM	4	1990
ZAM	0	1980
...

(8)- Từ table three của câu 3, viết lệnh truy vấn để có kết quả như minh họa sau:

• Table three

nationality	Height_200	Height_210	Height_220	Height_230
ZAM	1	0	0	0
VEN	8	2	0	0
UZB	2	0	0	0
USA	45	5	1	0
UKR	6	1	0	0
TUR	1	0	0	0
TUN	5	0	0	0
TTO	2	0	0	0
...

• *Kết quả thứ 1 cần thực hiện*

nationality	Quantity	HeightInCm
ZAM	1	Height_200
ZAM	0	Height_210
ZAM	0	Height_220
ZAM	0	Height_230
VEN	8	Height_200
VEN	2	Height_210
VEN	0	Height_220
VEN	0	Height_230
...

• *Kết quả thứ 2 cần thực hiện*

nationality	LowerThan220	GreaterThanOrEqual220	Height
ZAM	1	0	<220
ZAM	0	0	>=220
VEN	8	2	<220
VEN	0	0	>=220

(9)- Từ table four của câu 4, viết lệnh truy vấn để có kết quả như minh họa sau:

• *Table four*

sport	HeightInCm	Weight_100	Weight_110	Weight_120	Weight_130	Weight_140
aquatics	200.0	29	16	5	1	0
athletics	200.0	5	6	10	6	7
athletics	210.0	0	0	3	0	0
basketball	200.0	28	15	2	0	0
basketball	210.0	9	13	16	2	1
basketball	220.0	1	0	1	0	0
canoe	200.0	2	0	1	0	0
handball	210.0	0	2	2	0	0
volleyball	210.0	11	3	1	0	0

• *Kết quả cần thực hiện (sau khi đã loại bỏ các giá trị =0 trong table four và sắp xếp giảm dần theo Quantity)*

sport	HeightInCm	Quantity	Weight
rowing	200	45	Weight_100
aquatics	200	29	Weight_100
volleyball	200	29	Weight_100
basketball	200	28	Weight_100
handball	200	24	Weight_100
handball	200	23	Weight_110
aquatics	200	16	Weight_110
...

(Kết quả gồm 57 records)

(10)- Từ table five của câu 5, viết lệnh truy vấn để có kết quả như minh họa sau:
trong đó:

- Nếu là 2 cột weight_60 và weight_70 sẽ nhận chuỗi "*Lest than Or Equal 80*"
- Các cột còn lại sẽ nhận chuỗi "*Greater than 80*"

• *Table five*

HeightInCm	Weight_60	Weight_70	Weight_80	Weight_90	Weight_100
160	6	1	1	0	0
170	14	6	0	0	0
180	4	25	5	4	1
190	0	4	2	6	0
200	0	0	0	0	0

• *Kết quả cần thực hiện*

HeightInCm	Weight_100	Below80	Over80	Weight
160	0	6	1	LessThanOrEqual80
160	0	1	0	GreaterThan80
170	0	14	6	LessThanOrEqual80
170	0	0	0	GreaterThan80
180	1	4	25	LessThanOrEqual80
180	1	5	4	GreaterThan80

6.2. Data Science Salary

6.2.1. Pivot

(11)- Tạo thống kê số lượng CompanySize theo từng quốc gia (dựa trên Company Location). Yêu cầu sắp xếp kết quả giảm dần theo tên quốc gia.

CompanyLocation	Quantity S	Quantity M	Quantity L
Viet Nam	0	0	1
United States of America	54	2723	263
United Kingdom of Great Britain and Northern Ireland	8	144	20
United Arab Emirates	2	0	1
Ukraine	1	2	1
...

(12)- Tạo thống kê số lượng người có thu nhập (sau khi làm tròn SalaryInUsd đến phần ngàn) theo từng quốc gia (dựa trên Company Location). Yêu cầu:

- Sắp xếp kết quả giảm dần theo tên quốc gia.
- Mặc dù có nhiều mức thu nhập nhưng kết quả chỉ cần thể hiện kết quả của các mức 10.000, 20.000, 30.000, 90.000, 100.000
- Kết quả chỉ hiện ra những quốc gia có cột Over_10000>0.

name	Over_10000	Over_20000	Over_30000	Over_90000	Over_100000
Viet Nam	1	0	0	0	0
United States of America	4	4	8	113	178
Ukraine	1	0	0	0	0
Turkey	2	1	2	0	0
Switzerland	1	0	0	0	1
Viet Nam	1	0	0	0	0

(13)- Thống kê số lượng nhân viên tham gia theo từng loại việc làm (*Employment Type*) đối với theo từng chức danh nghề nghiệp (*JobTitleName*). Yêu cầu sắp xếp kết quả tăng dần theo *JobTitleName*.

JobTitleName	NumberOf_Full-time	NumberOf_Consultants	NumberOf_Freelancer	NumberOf_Part-time employees
3D Computer Vision Researcher	3	0	0	1
AI Developer	11	0	0	0
AI Programmer	2	0	0	0
AI Scientist	14	0	0	2
Analytics Engineer	102	1	0	0

6.2.2. UnPivot

(14)- Từ table eleven của câu 11, viết lệnh truy vấn để có kết quả như minh họa sau:

- *Table eleven*

CompanyLocation	Quantity_S	Quantity_M	Quantity_L
Viet Nam	0	0	1
United States of America	54	2723	263
United Kingdom of Great Britain and Northern Ireland	8	144	20
United Arab Emirates	2	0	1
Ukraine	1	2	1
...

- *Kết quả cần thực hiện*

<i>name</i>	<i>quantity</i>	<i>EmploymentType</i>
Albania	1	Quantity_S
Albania	0	Quantity_M
Albania	0	Quantity_L
Algeria	0	Quantity_S
Algeria	1	Quantity_M
Algeria	0	Quantity_L
...

(15)- Từ table twelve của câu 12, viết lệnh truy vấn để có kết quả như minh họa sau:

- *Table twelve*

name	Over_10000	Over_20000	Over_30000	Over_90000	Over_100000
Viet Nam	1	0	0	0	0
United States of America	4	4	8	113	178
Ukraine	1	0	0	0	0
Turkey	2	1	2	0	0
Switzerland	1	0	0	0	1
Viet Nam	1	0	0	0	0

- *Kết quả cần thực hiện*

<i>name</i>	<i>quantity</i>	<i>EmploymentType</i>
Albania	1	Quantity_S
Albania	0	Quantity_M
Albania	0	Quantity_L
Algeria	0	Quantity_S
Algeria	1	Quantity_M
Algeria	0	Quantity_L
...

(16)- Từ table thirteen của câu 13, viết lệnh truy vấn để có kết quả như minh họa sau:

- *Table thirteen*

<i>JobTitleName</i>	<i>NumberOf_Full-time</i>	<i>NumberOf_Consultants</i>	<i>NumberOf_Freelancer</i>	<i>NumberOf_Part-time employees</i>
3D Computer Vision Researcher	3	0	0	1
AI Developer	11	0	0	0
AI Programmer	2	0	0	0
AI Scientist	14	0	0	2
Analytics Engineer	102	1	0	0

• *Kết quả cần thực hiện*

<i>JobTitleName</i>	<i>quantity</i>	<i>EmploymentType</i>
3D Computer Vision Researcher	3	NumberOf Full-time
3D Computer Vision Researcher	0	NumberOf Consultants
3D Computer Vision Researcher	0	NumberOf Freelancer
3D Computer Vision Researcher	1	NumberOf Part-time employees
AI Developer	11	NumberOf Full-time
AI Developer	0	NumberOf Consultants
AI Developer	0	NumberOf Freelancer
AI Developer	0	NumberOf Part-time employees
...

BÀI TẬP SỐ 7

(GROUP BY - RollUp)

7.1. Athletes

1. Tạo thống kê như minh họa sau:

<i>nationality</i>	<i>sex</i>	<i>GoldMedal</i>	<i>SilverMedal</i>	<i>BronzeMedal</i>
<i>null</i>	<i>null</i>	666	655	704
AFG	<i>null</i>	0	0	0
AFG	male	0	0	0
AFG	female	0	0	0
ALB	<i>null</i>	0	0	0
ALB	female	0	0	0
ALB	male	0	0	0
ALG	<i>null</i>	0	2	0
ALG	male	0	2	0
ALG	female	0	0	0
AND	<i>null</i>	0	0	0
AND	female	0	0	0
AND	male	0	0	0
...	...			

2. Thống kê tương tự như trên nhưng loại đi những dòng không có bất kỳ huy chương nào trong 3 loại gold, silver và bronze. Lưu ý, những nước mà phái nam và nữ đều không có huy chương sẽ bị loại bỏ trong kết quả (như 2 nước AFG và ALB). Kết quả như minh họa sau:

<i>nationality</i>	<i>sex</i>	<i>GoldMedal</i>	<i>SilverMedal</i>	<i>BronzeMedal</i>
<i>null</i>	<i>null</i>	666	655	704
ALG	<i>null</i>	0	2	0
ALG	male	0	2	0
ARG	<i>null</i>	21	1	0
ARG	male	19	1	0
ARG	female	2	0	0
ARM	<i>null</i>	1	3	0
ARM	male	1	3	0
AUS	<i>null</i>	23	34	25
AUS	female	20	14	4
AUS	male	3	20	21
...		

3. Thống kê số lượng từng loại huy chương theo giới tính và lứa tuổi. Kết quả như minh họa sau:

<i>sex</i>	<i>age</i>	<i>GoldMedal</i>	<i>SilverMedal</i>	<i>BronzeMedal</i>
<i>null</i>	<i>null</i>	666	655	704
female	<i>null</i>	318	317	331
male	<i>null</i>	348	338	373
female	14	0	0	0
male	14	0	0	0
female	15	1	0	0
male	16	0	0	0
female	16	2	3	5
male	17	0	0	0
female	17	1	1	5
...		

4. Thống kê số lượng từng loại huy chương theo sex và sport cho những người có age=53. Kết quả như minh họa sau:

sex	sport	GoldMedal	SilverMedal	BronzeMedal
null	null	1	1	3
female	null	0	1	0
female	shooting	0	0	0
female	equestrian	0	1	0
female	table tennis	0	0	0
male	null	1	0	3
male	shooting	0	0	1
male	equestrian	1	0	2

7.2. Data Science Salary

5. Thống kê thu nhập bình quân theo quốc gia nơi đặt văn phòng chính của công ty (Company Location), size của công ty, các chức danh và lương bình quân tính theo USD.

NAME	SIZE	jobtitlename	AVG_SalaryInUsd
Viet Nam	null	null	12000.0
Viet Nam	Large	null	12000.0
Viet Nam	Large	Data Engineer	12000.0
United States of America	null	null	151822.009539474
United States of America	Medium	null	152495.08777084091
United States of America	Medium	Research Engineer	168156.5806451613
United States of America	Medium	BI Developer	129833.33333333334
United States of America	Medium	Data Engineer	153372.28789659232
United States of America	Medium	Data Scientist	160351.61475409844
United States of America	Medium	Machine Learning Engineer	175709.03108808296
United States of America	Large	null	154474.81368821289
United States of America	Large	Machine Learning Engineer	176500.0
...

6. Thống kê số lượng người theo các quốc gia thuộc khu vực Châu Á (region='Asia') về nơi đặt văn phòng chính của công ty (Company Location), hình thức làm việc (EmploymentType), chức danh nghề nghiệp (JobTitleName) và số lượng người trong từng nhóm.

name	EmploymentTypeName	JobTitleName	NumberOfEmployees
Viet Nam	null	null	1
Viet Nam	Full-time	null	1
Viet Nam	Full-time	Data Engineer	1
United Arab Emirates	null	null	3
United Arab Emirates	Full-time	null	3
United Arab Emirates	Full-time	Lead Data Scientist	1
United Arab Emirates	Full-time	Machine Learning Engineer	2
Turkey	null	null	5
Turkey	Full-time	null	5

Turkey	Full-time	Data Scientist	3
Turkey	Full-time	Data Engineer	2
...

7. Tạo thống kê để có kết quả như hình minh họa sau:

region	CompanySize	EmploymentTypeName	JobTitleName	CountOf Employees	Avg_ Salary	Min_ Salary	Max_ Salary
null	null	null	null	3755	137570	5132	450000
Africa	null	null	null	14	53306	7000	200000
Africa	L	null	null	3	60000	30000	100000
Africa	L	Full-time	null	3	60000	30000	100000
Africa	L	Full-time	Compliance Data Analyst	1	30000	30000	30000
Africa	L	Full-time	Data Scientist	1	50000	50000	50000
Africa	L	Full-time	Data Quality Analyst	1	100000	100000	100000
Africa	M	null	null	4	55005	22800	100000
Africa	M	Full-time	null	3	40006	22800	54685
Africa	M	Full-time	Data Analyst	2	48609	42533	54685
Africa	M	Part-time employees		1	100000	100000	100000
Africa	M	Part-time employees	Data Scientist	1	100000	100000	100000
Africa	M	Full-time	Machine Learning Scientist	1	22800	22800	22800
Africa	S	null	null	7	49467	7000	200000
Africa	S	Full-time	null	7	49467	7000	200000
Africa	S	Full-time	AI Scientist	1	30000	30000	30000
Africa	S	Full-time	Autonomous Vehicle Technician	1	7000	7000	7000
Africa	S	Full-time	BI Data Analyst	1	9272	9272	9272
Africa	S	Full-time	Applied Data Scientist	1	80000	80000	80000
Africa	S	Full-time	Machine Learning Software Engineer	1	10000	10000	10000
Africa	S	Full-time	Data Analyst	1	10000	10000	10000
Africa	S	Full-time	BI Analyst	1	200000	200000	200000
Americas	null	null	null	3168	150178	5679	450000
Americas	L	null	null	290	148025	6270	423000
Americas	L	Full-time	null	285	149123	6270	423000
Americas	L	Full-time	Data Analyst	26	87464	7799	202800

BÀI TẬP SỐ 8 (Optimized Queries)

8.1. Data Science Salary

1. Cho biết có phải năm 2022 là năm có lương cao nhất theo từng năm của những người được trả bằng đơn vị tiền tệ là 'United Kingdom Pound'. Trả lời có hoặc không.
2. Cho biết tên những tiểu vùng (*sub-region*) nào mà chức danh (*JobTitleName*) 'Data Analyst' có mức lương (*SalaryInUsd*) làm việc toàn thời gian (*EmploymentTypeName* = 'Full-time') nằm trong khoảng từ 80.000 đến 100.000.
3. Cho biết tên những chức danh nghề nghiệp (*JobTitleName*) nào có văn phòng chính ở Châu Âu (*region* = 'Europe') trả lương (*SalaryInUsd*) dưới 80.000 cho những người mà bản địa của họ (*Employee Residence*) không phải là nước Mỹ ('United States of America').
4. Cho biết chức tên những chức danh nghề nghiệp (*JobTitleName*) nào mà số lượng người thuộc chức danh đó không nằm trong 20 chức danh có số lượng người là đông nhất và cũng không nằm trong 20 chức danh có số lượng người là thấp nhất.
5. Cho biết tên những nước (dựa trên *CompanyLocation*) có trên 70% chức danh nghề nghiệp có trong table *JobTitle* hoặc những nước không thuộc châu Mỹ (*region* <> 'Americas') có thanh toán lương (*SalaryCurrency*) không phải là USD.
6. Cho biết tên chức danh nghề nghiệp (*JobTitleName*) có *SalaryInUsd* nằm trong khoảng từ 30.000 đến 90.000 tại những nước mà tên nước (*name*) ký tự đầu và ký tự cuối là giống nhau (không phân biệt hoa thường) thuộc *region* cũng có ký tự đầu và ký tự cuối là giống nhau (không phân biệt hoa thường) và mã đơn vị tiền tệ (thuộc tính *Symbol* trong table *CurrencyCode* có giá trị là ký hiệu dollar (\$).

8.2. Sales

7. Cho biết tên những loại hàng (*CategoryName*) mà khách hàng nữ người Australia trong độ tuổi từ 45-50 thường mua và ký tự thứ 4 của tên sản phẩm là ký tự 'e' hoặc 'n'.
8. Cho biết tên những sản phẩm có chứa 2 ký tự 'o' (2 ký tự này không phân biệt ký tự hoa/thường và có thể đặt liền nhau hoặc tách rời) thuộc loại hàng là 'Bikes' mà khách hàng người Pháp (France) dưới 30 tuổi chưa từng mua.

9. Cho biết tên những tiểu bang/tỉnh/thành phố nào mà nước của tiểu bang/tỉnh/thành phố đó có chứa 3 ký tự A (không phân biệt ký tự hoa/thường, ví dụ Australia, Canada, ...) và khách hàng nữ có tuổi từ 35-45 thuộc tiểu bang đó đã mua những sản phẩm mà tên sản phẩm có chứa dấu phẩy (,).
 10. Cho biết tuổi thấp nhất và tuổi cao nhất của những khách hàng nam, người Đức (Germany) thuộc 1 trong 2 bang “*Saarland*” và “*Hamburg*” đã mua sản phẩm thuộc 1 trong các nhóm hàng mà tên của nhóm hàng có ký tự đầu và cuối giống nhau (không phân biệt ký tự hoa/thường, ví dụ: Shorts, Socks, ...).
 11. Cho biết khách hàng người nước Anh (United Kingdom) thuộc phái nào (kết quả trả lời là “male” hoặc “female”) đã mua những sản phẩm có giá (Unit_Cost) từ 7 đến 12 với số lượng (Order_Quantity) dưới 40 hoặc trên 100 và tên sản phẩm có chứa 2 ký tự “0” và “a” sao cho ký tự nào đứng trước cũng được.
-

BÀI TẬP SỐ 9 (Struct - Array & Record)

9.1. MyFriend

Cho schema của table People có nested column và repeated field như sau:

```
[
  {
    "name": "id",
    "type": "STRING",
    "mode": "NULLABLE"
  },
  {
    "name": "first_name",
    "type": "STRING",
    "mode": "NULLABLE"
  },
  {
    "name": "last_name",
    "type": "STRING",
    "mode": "NULLABLE"
  },
  {
    "name": "dob",
    "type": "DATE",
    "mode": "NULLABLE"
  },
  {
    "name": "addresses",
    "type": "RECORD",
    "mode": "REPEATED",
    "fields": [
      {
        "name": "status",
        "type": "STRING",
        "mode": "NULLABLE"
      },
      {
        "name": "address",
        "type": "STRING",
        "mode": "NULLABLE"
      },
      {
        "name": "city",
        "type": "STRING",
        "mode": "NULLABLE"
      },
      {
        "name": "state",
        "type": "STRING",
        "mode": "NULLABLE"
      },
      {
        "name": "zip",
        "type": "STRING",
        "mode": "NULLABLE"
      },
      {
        "name": "numberOfYears",
        "type": "STRING",
        "mode": "NULLABLE"
      }
    ]
  }
]
```

Minh họa dữ liệu sẽ được lưu trong table

id	first_name	last_name	dob	Nested data					
				addresses					
				status	address	city	state	zip	numberOfYears
1	John	Doe	1968-01-22	current	123 First Avenue	Seattle	WA	11111	1
				previous	456 Main Street	Portland	OR	22222	5
2	Jane	Doe	1980-10-16	current	789 Any Avenue	New York	NY	33333	2
				previous	321 Main Street	Hoboken	NJ	44444	3
...

Yêu cầu: Tạo table bằng 3 cách:

- Tạo table MyFriend_01 bằng cách thủ công
- Tạo table MyFriend_02 dựa trên nội dung file json do GV cung cấp.
- Viết lệnh tạo table MyFriend_03 cùng dữ liệu như trong hình minh họa ở trên.

9.2. Employees_Departments

- Tạo dataset Employees_Departments
- Từ dữ liệu do GV cung cấp, học viên tạo table:
 - Departments từ file Departments.csv
 - Employees từ nội dung truy vấn có trong file Employees.txt.

(Xem mô tả và minh họa dữ liệu trong phần “Giới thiệu các CSDL dùng trong tài liệu”).

- Viết các truy vấn sau:

- Hiện thị toàn bộ thông tin có trong table Employees, sắp xếp tăng dần theo cột dob.
- Cho biết còn bao nhiêu nhân viên đang còn đi làm (nhắc lại: khi trên field của 1 nhân viên nào đó không có giá trị NULL, tức là nhân viên đã nghỉ việc.
- Cho biết mã số, họ, tên, số điện thoại và email của những nhân viên đã nghỉ việc. Kết quả có dạng như minh họa sau:

EmpID	name	fName	Phone	Email
3	Emma	Brook	290807369	fhfdg@gmail.com

- Cho biết tên phòng ban cùng mã số, họ, tên của tất cả các nhân viên đang công tác tại phòng ban đó. Sắp xếp kết quả theo tên phòng ban. Kết quả có dạng như hình sau

DepartmentName	EmpID	name	fname
Finance	14	Florence	Chandler
Finance	9	Bertha	Lupe
Finance	11	Annie	Prentice
Finance	1	Mary	Arnulfo
Graphic Designing	12	Clara	Benson
Graphic Designing	18	Nellie	Emmitt

Graphic Designing	7	Ida	Dwane
Graphic Designing	19	Grace	Isidro
Graphic Designing	6	Margaret	Darold
Graphic Designing	13	Ella	Brandt
Public Relations (PR)	4	Elizabeth	Damien
Public Relations (PR)	16	Martha	Deborah
Public Relations (PR)	8	Alice	Kalvin
Public Relations (PR)	15	Cora	Christoper
Public Relations (PR)	2	Anna	Arron
Research & Development (R&D)	10	Sarah	Mel
Research & Development (R&D)	17	Laura	Elroy
Research & Development (R&D)	5	Minnie	Darian
Research & Development (R&D)	20	Carrie	Kennth

(v).- Cho biết tên phòng ban cùng số lượng nhân viên hiện có Kết quả có dạng như hình sau:

<i>departmentName</i>	<i>NumberOfEmp</i>
Finance	4
Graphic Designing	6
Management	0
Public Relations (PR)	5
Research & Development (R&D)	4

(vi).- Cho biết mã số, họ, tên, năm bắt đầu làm việc (*FirstDayOfWork*), Số năm thâm niên công tác (*YearsOfService*). Kết quả có dạng như hình sau:

<i>EmpID</i>	<i>name</i>	<i>fName</i>	<i>FirstDayOfWork</i>	<i>YearsOfService</i>
1	Mary	Arnulfo	2010	13
2	Anna	Arron	2010	13
4	Elizabeth	Damien	2010	13
5	Minnie	Darian	2011	12
6	Margaret	Darold	2015	8
7	Ida	Dwane	2011	12
8	Alice	Kalvin	2010	13
9	Bertha	Lupe	2010	13
10	Sarah	Mel	2010	13
11	Annie	Prentice	2018	5
12	Clara	Benson	2011	12
13	Ella	Brandt	2010	13
14	Florence	Chandler	2011	12
15	Cora	Christoper	2010	13
16	Martha	Deborah	2015	8
17	Laura	Elroy	2010	13
18	Nellie	Emmitt	2015	8
19	Grace	Isidro	2010	13
20	Carrie	Kennth	2010	13

(vii).- Cho biết mã số, họ, tên của những nhân viên có **số tháng** thâm niên công tác (*YearsOfService*) nhiều nhất. Kết quả có dạng như hình sau:

<i>EmpID</i>	<i>name</i>	<i>fName</i>	<i>MonthsOfService</i>
1	Mary	Arnulfo	165
4	Elizabeth	Damien	165
8	Alice	Kalvin	165
9	Bertha	Lupe	165
13	Ella	Brandt	165
17	Laura	Elroy	165
20	Carrie	Kennth	165

BÀI TẬP TỔNG HỢP

10.1. Sales

I. Tạo cơ sở dữ liệu Cung cấp hàng hóa.

1.- Tạo các table từ các file CSV do GV cung cấp. Do các file CSV này không có tên field nên học viên cần viết lệnh để đổi tên cho các field như sau:

(i).- **Table States**

- string_field_0 TO CountryCode
- string_field_1 TO StateName

(ii).- **Table Categories**

- string_field_0 TO Code
- string_field_1 TO Category

(iii).- **Table Countries**

- string_field_0 TO CountryCode
- string_field_1 TO Name

(iv).- **Table Sub_Categories**

- int64_field_0 TO Sub_CategoryCode
- string_field_1 TO Sub_Category
- string_field_2 TO CategoryCode

(v).- **Table Products**

- string_field_0 TO Sub_CategoryCode
- string_field_0 TO ProductName

(vi).- **Table Data**

- int64_field_0 TO Customer_Age
- string_field_1 TO Customer_Gender
- string_field_2 TO StateName
- string_field_3 TO Product
- int64_field_4 TO Order_Quantity
- int64_field_5 TO Unit_Cost
- int64_field_6 TO Cost
- int64_field_7 TO Profit
- int64_field_8 TO Revenue

2.-Viết lệnh tạo table cùng các ràng buộc khóa chính (primary key), khóa ngoại (forfeign key).

II. Viết các lệnh truy vấn sau:

1.- Cho biết tên những tiểu bang/tỉnh/thành phố nào mà ký tự đầu và ký tự cuối là giống nhau (không phân biệt ký tự hoa/thường). Ví dụ: Alberta, Nordrhein-Westfalen, Ontario, ...

2.- Cho biết tên những nhóm hàng (Sub_CategoryName) mang lại lợi nhuận (Profit) nhiều hoặc ít nhất theo mẫu sau, với yêu cầu lần lượt thực hiện bằng 3 cách sau:

- Sử dụng hàm ANY_VALUE
- Sử dụng hàm MAX_BY/MIN_BY

- Không sử dụng các hàm ANY_VALUE và MAX_BY/MIN_BY.

Kết quả truy vấn:

Có sử dụng các hàm ANY_VALUE và MAX_BY/MIN_BY		Không sử dụng các hàm ANY_VALUE và MAX_BY/MIN_BY	
max_profit	min_profit	max_profit	min_profit
Touring Bikes	Jerseys	null	Jerseys
		Touring Bikes	null
		Vests	null

- 3.- Tên những quốc gia mà khách hàng không mua sản phẩm 'LL Road Tire' với số lượng (Order_Quantity) là 32.

Kết quả truy vấn:

Row	Name
1	Canada
2	Germany
3	United Kingdom

- 4.- Cho biết khách thuộc tiểu bang/tỉnh/thành phố nào? Lứa tuổi nào? Giới tính nào? Là đông (nhiều) nhất.

Kết quả truy vấn:

Row	State	Customer_Age	Customer_Gender	NumberOfCustomer
1	California	34	M	534

- 5.- Viết truy vấn để có kết quả như minh họa sau:

Kết quả truy vấn:

Row	Name	StateName	Average_Profit	MaxProfitOfCountry	MinProfitOfCountry
1	Australia	Victoria	237.0	310.0	235.0
2	Australia	South Australia	272.0	310.0	235.0
3	Australia	New South Wales	301.0	310.0	235.0
4	Australia	Queensland	310.0	310.0	235.0
5	Australia	Tasmania	235.0	310.0	235.0
6	Canada	British Columbia	262.0	417.0	100.0
7	Canada	Alberta	417.0	417.0	100.0
8	Canada	Ontario	100.0	417.0	100.0
9	France	Seine (Paris)	220.0	396.0	170.0

- 6.- Viết truy vấn sau sao cho tối ưu hóa, với yêu cầu: cho biết tên sản phẩm mà khách hàng nữ người Mỹ (CountryName= 'United States') ở lứa tuổi từ 40 đến 50 đã mua sản phẩm thuộc loại 'Clothing' (CategoryName='Clothing') và đem lại lợi nhuận (Profit) trên 100.

Kết quả truy vấn:

Row	Product
1	Fender Set - Mountain
2	Long-Sleeve Logo Jersey, L
3	Long-Sleeve Logo Jersey, M
4	Long-Sleeve Logo Jersey, S
5	Women's Mountain Shorts, S
6	Women's Mountain Shorts, M
7	Half-Finger Gloves, S
8	Half-Finger Gloves, L

Results per page: 50 1 - 21 of 21

- 7.- Giả sử người ta phân tuổi của khách hàng thành 3 nhóm: dưới 25, từ 25 đến 50 và trên 50. Yêu cầu thống kê số lượng khách của từng quốc gia theo lứa tuổi theo minh họa sau:

Kết quả truy vấn:

Name ▾	Under25 ▾	Between25And50 ▾	Over50 ▾
Australia	4382	17520	2034
France	1590	8274	1134
United States	5698	28020	5488
United Kingdom	2124	10056	1440
Canada	2326	10032	1820
Germany	1708	8278	1112

- 8.- Tương tự như trên, giả sử người ta phân tuổi của khách hàng thành 3 nhóm: dưới 25, từ 25 đến 50 và trên 50. Sử dụng mệnh đề FROM với PIVOT để tạo thống kê sau. Biết rằng thống kê này chỉ tính riêng cho các State thuộc nước Úc (Name='Australia')

Kết quả truy vấn:

StateName ▾	<25 years old ▾	>50 years old ▾	25-50 years old ▾
Victoria	1142	538	4336
South Australia	334	86	1144
New South Wales	1862	964	7586
Queensland	886	394	3940
Tasmania	158	52	514

- 9.- Tạo thống kê theo mẫu sau

Kết quả truy vấn:

Name ▾	StateName ▾	Product ▾	Sum_Revenue ▾
null	null	null	85271008
Australia	null	null	21302059
Australia	New South Wales	null	9203495
Australia	New South Wales	LL Road Tire	23732
Australia	New South Wales	Fender Set - Mountain	77884
Australia	New South Wales	Classic Vest, S	27114
Australia	New South Wales	Patch Kit/8 Patches	22503
Australia	New South Wales	Road Tire Tube	15110
Australia	New South Wales	Water Bottle - 30 oz.	77787
Australia	New South Wales	Mountain Tire Tube	48078
Australia	New South Wales	Touring Tire Tube	20983
Australia	New South Wales	Long-Sleeve Logo Jersey, XL	37729
Australia	New South Wales	Long-Sleeve Logo Jersey, M	52178
Australia	New South Wales	Long-Sleeve Logo Jersey, S	11140
Australia	New South Wales	Bike Wash - Dissolver	22455

- 10.- Cho biết tên nước mà khách hàng nam với tuổi nhỏ hơn 24 đã đặt mua sản phẩm thuộc loại 'Accessories' (Category='Accessories') và đem lại lợi nhuận trên 1000.

Kết quả truy vấn:

Row ▾	Name ▾
1	Germany

11.- Tạo truy vấn để có kết quả như hình minh họa, biết rằng:

- Kết quả chỉ bao gồm những mã loại hàng (Sub_CategoryCode) là 'A'.
- Các tên cột được tính và chuyển đổi dựa trên hàm COUNT.
- Dữ liệu thống kê trong bảng được tính dựa trên việc tính tổng của field lợi nhuận (Profit).

Kết quả truy vấn:

Sub_CategoryName ▾	<5000 ▾	<10000 ▾	<20000 ▾	>=20000 ▾
Fenders	null	865049	null	null
Tires and Tubes	null	null	null	2727713
Bottles and Cages	null	null	810606	null
Cleaners	null	194016	null	null
Helmets	null	null	3381178	null
Hydration Packs	null	1713023	null	null
Bike Racks	305326	null	null	null
Bike Stands	506390	null	null	null

12.- Thống kê 3 mặt hàng có số lượng bán là nhiều nhất ở mỗi nước. Kết quả có dạng như hình sau. Lưu ý: nếu tại 1 nước bất kỳ có 4 sản phẩm cùng nằm trong top 3” thì cũng chỉ cho xuất hiện tối đa 3 sản phẩm

Row ▴	Name ▾	Ranking ▾	ProductName ▾	Sum_Quantity ▾
1	Australia	1	Water Bottle - 30 oz.	65528
2	Australia	2	Patch Kit/8 Patches	32668
3	Australia	3	AWC Logo Cap	26660
4	Canada	1	Water Bottle - 30 oz.	45880
5	Canada	2	Mountain Tire Tube	27546
6	Canada	3	Patch Kit/8 Patches	25926
7	France	1	Water Bottle - 30 oz.	29762
8	France	2	Patch Kit/8 Patches	16973
9	France	3	AWC Logo Cap	14380
10	Germany	1	Water Bottle - 30 oz.	25152
11	Germany	2	AWC Logo Cap	16810
12	Germany	3	Sport-100 Helmet, Black	14242
13	United Kingdom	1	Water Bottle - 30 oz.	39502
14	United Kingdom	2	AWC Logo Cap	20508
15	United Kingdom	3	Touring Tire Tube	14998
16	United States	1	Water Bottle - 30 oz.	122348
17	United States	2	Patch Kit/8 Patches	56408
18	United States	3	Fender Set - Mountain	56062

13.- Cho biết tên những bang (*state*) mà nam giới ở bang đó chưa từng bao giờ mua hàng (không có trong dữ liệu)

Minh họa kết quả truy vấn (gồm tên 8 bang):

StateName
Alabama
Massachusetts
Mississippi
Missouri
Montana
North Carolina
Ontario
South Carolina

14.- Tính tỷ lệ số lượng khách hàng nữ/số lượng khách hàng nam theo từng bang. Nếu mẫu số của tỷ lệ này là 0, sẽ in ra “Nam giới không mua hàng”. Lưu ý kết quả có kèm dấu phần trăm (%)

Minh họa kết quả truy vấn (gồm tên 53 bang):

StateName	Ratio
Yveline	.86%
Wyoming	.33%
Washington	.93%
Virginia	.00%
Victoria	.79%
Val de Marne	.76%
Val d'Oise	2.67%
Utah	.00%
Texas	6.50%
Tasmania	.65%
South Carolina	Nam giới không mua hàng
South Australia	.97%