Team K-BTS Project Submission

💪 Checklist

- [x] Project Proposal
- [x] A simple summary of open-source libraries comparison linked to project.
- [x] A list of ideas generated, consisting of conceptual sketches, function, and key components.
- [x] A simple written overview describing the database schema, table structure, and entity-relationship model.
- [x] The source code for your solutions to finish this task.
- [x] Final report that provides accurate answers to the questions, with charts (flexible format; can be a PowerPoint, document, etc.).

🗺️ Exploration + Solution

Q1. What are the regional sales in the best performing country?

Q2. What is the relationship between annual leave taken and bonus?

Q3. What is the relationship between Country and Revenue?

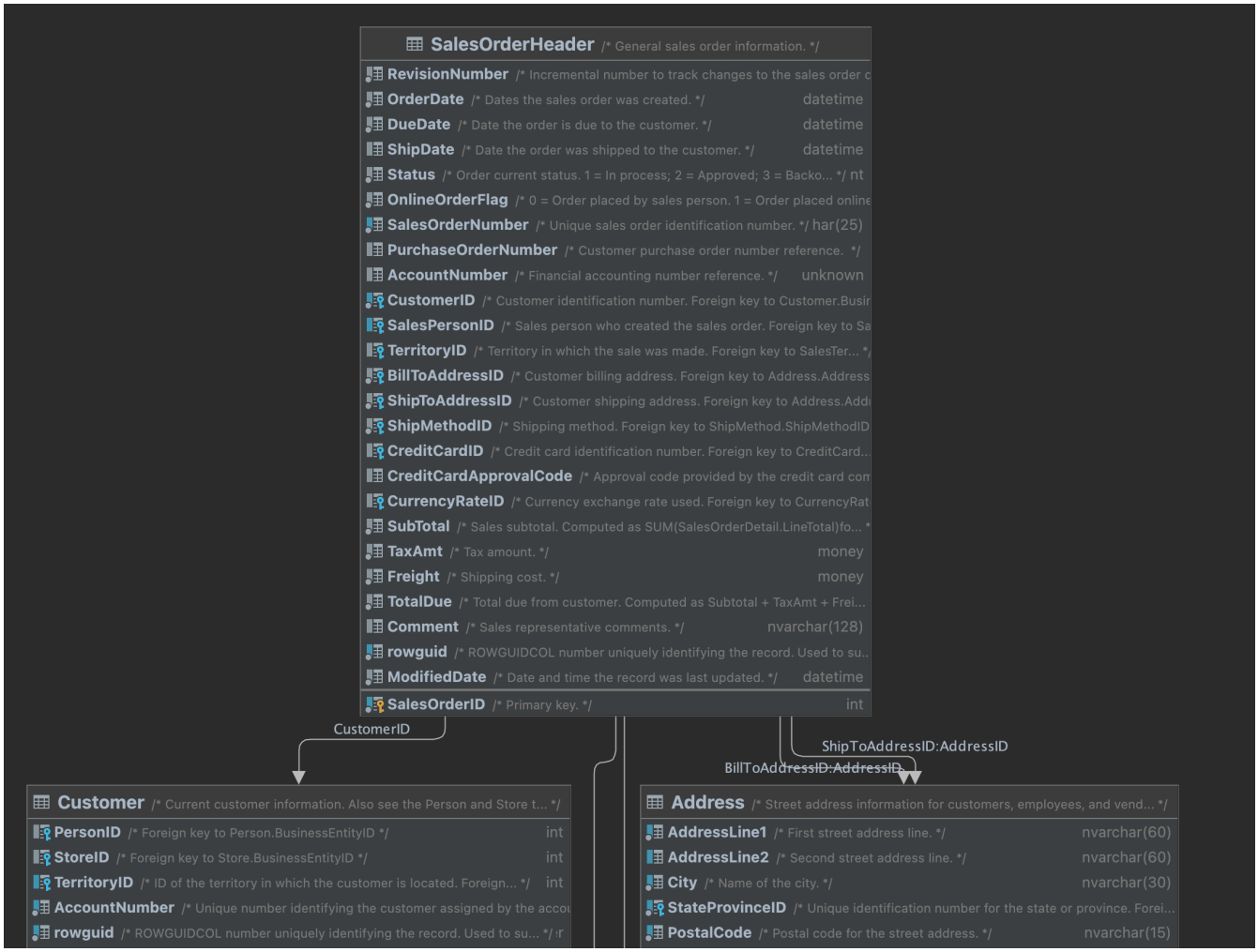
Q4. What is the relationship between sick leave and Job Title (PersonType)?

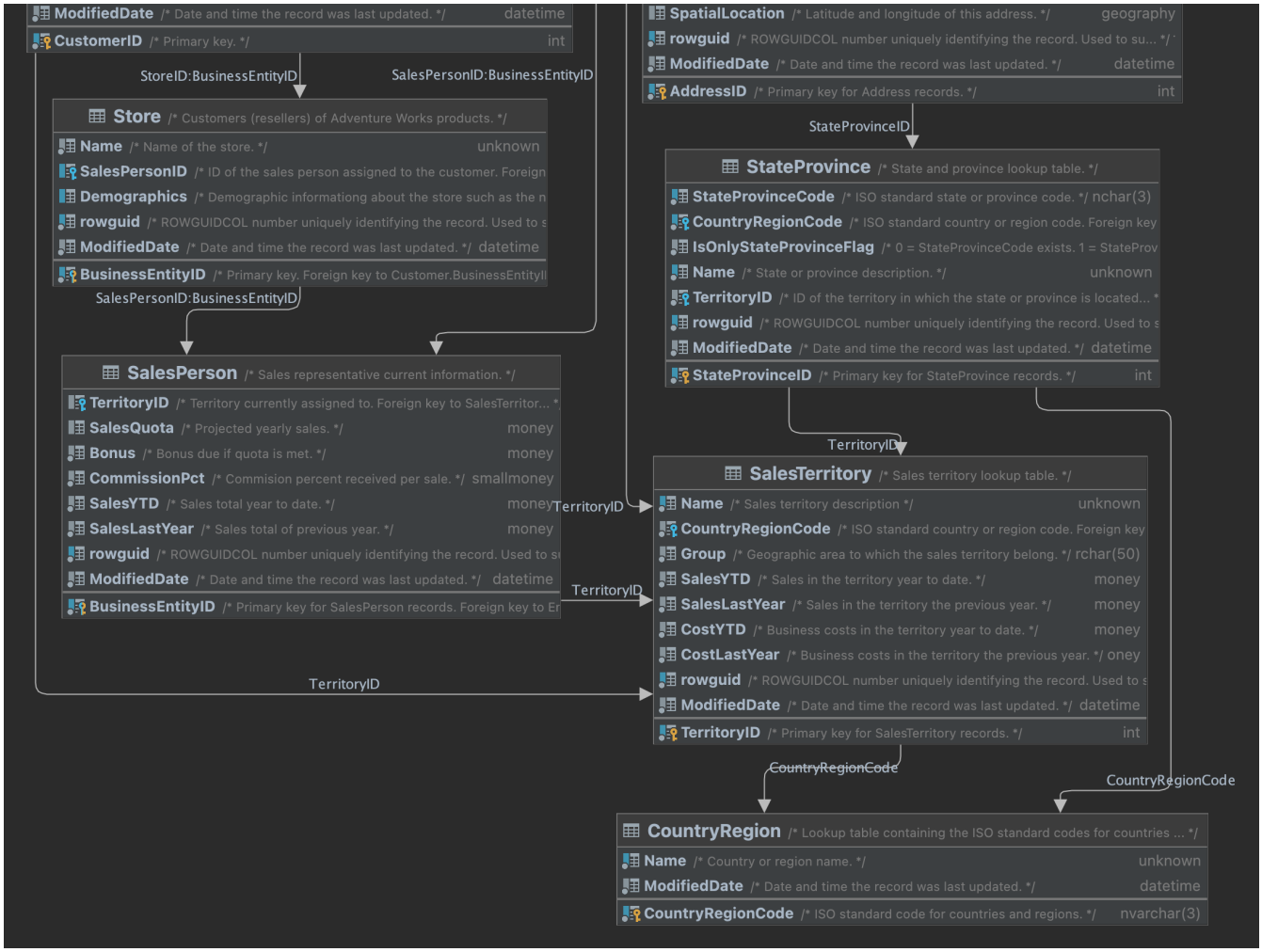
Q5. What is the relationship between store trading duration and revenue?

Q6. What is the relationship between the size of the stores, number of employees and revenue?

Q1. What are the regional sales in the best performing country?

Related Entity & Relationships





SQL

- We create a common table expression from a query containing the sums of the total due from the SalesOrderHeader that has transaction information and the corresponding region code.
- We then use this CTE on an inner join to match the country region code that has the highest sales.
- This makes sure we are selecting the regional information from the country with the highest sales in the query.

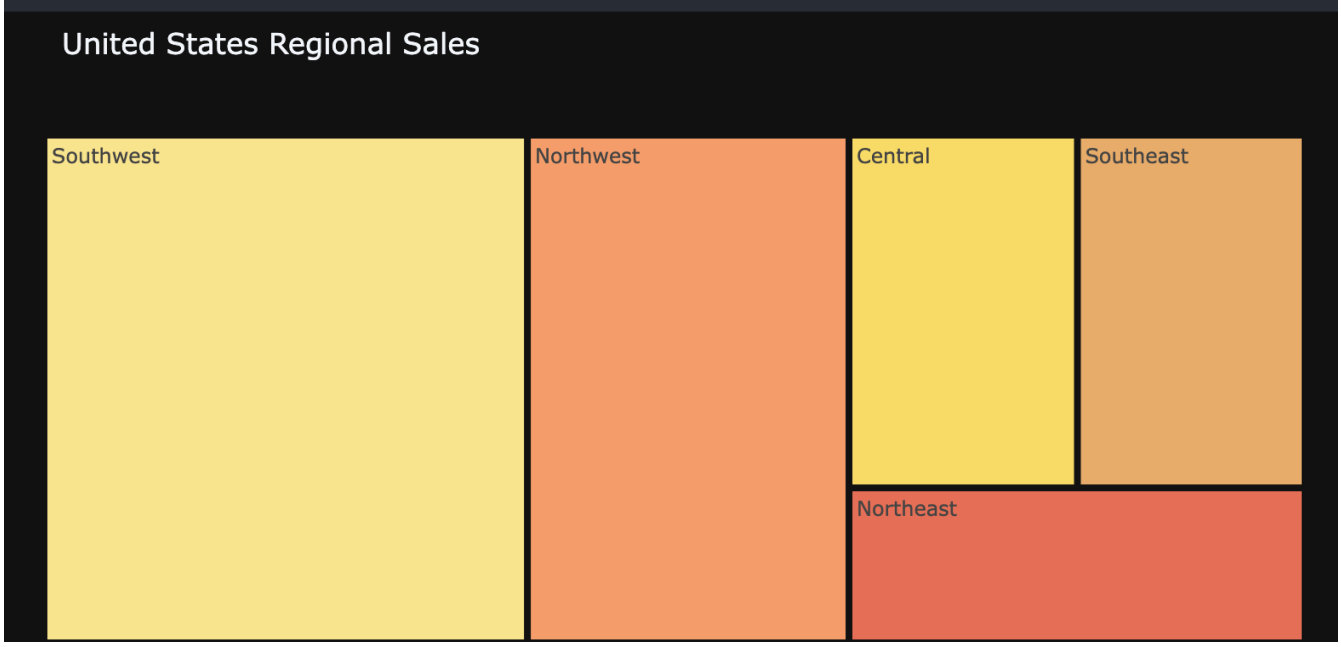
```
with MTSC as (select *
              from (SELECT SUM(soh.TotalDue) AS TotalSales, st.
CountryRegionCode AS CountryCode
                    FROM Sales.SalesOrderHeader AS soh
                             INNER JOIN Sales.Customer AS cu ON soh.
CustomerID = cu.CustomerID
                             INNER JOIN Sales.SalesTerritory AS st ON
cu.TerritoryID = st.TerritoryID
                    GROUP BY st.CountryRegionCode) as TSPC)

SELECT SUM(soh.TotalDue) AS TotalSales, st.TerritoryID, st.Name
FROM Sales.SalesOrderHeader AS soh
         INNER JOIN Sales.Customer AS cu ON (soh.CustomerID = cu.
CustomerID)
         INNER JOIN Sales.SalesTerritory AS st
                    ON (cu.TerritoryID = st.TerritoryID) AND st.
CountryRegionCode = (SELECT CountryCode

FROM MTSC

WHERE TotalSales = (SELECT MAX(TotalSales) FROM MTSC))
GROUP BY st.CountryRegionCode, st.TerritoryID, st.Name
ORDER BY TotalSales DESC
```

Assumptions

- All sales in SalesTerritory are converted to USD. Even if you don't assume this conversion the US is still extremely far ahead.
- We used SalesYTD to compare the regions sales amounts as this includes historical and current data.
- We compared Sales instead of revenue as it is more representative of business performance as revenue is controlled by external factors that are outside of scope of analysis. (eg: Rent, Taxes, ect)

Output & Visualisation

	TotalSales	TerritoryID	Name
0	2.715059e+07	4	Southwest
1	1.806166e+07	1	Northwest
2	8.913299e+06	3	Central
3	8.884099e+06	5	Southeast
4	7.820210e+06	2	Northeast

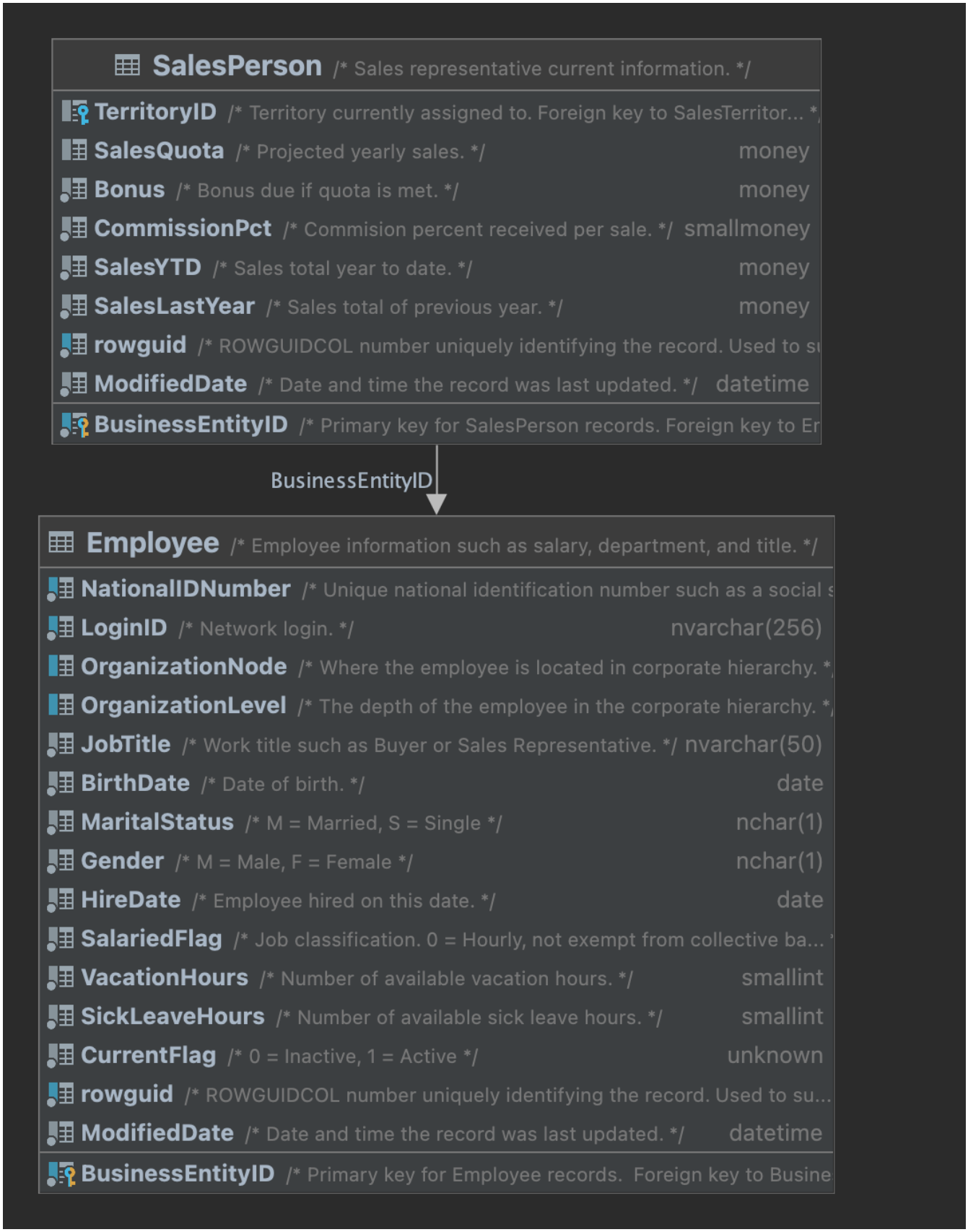


Answer & Further Insights

- The Southwest region has the largest share of sales, and is greater than the Central, Southeast and Northeast regions combined, or the Northwest and Northeast regions combined.
- The Central and Southeast regions are nearly identical in sales totals.
- Central and Southeast combined is similar to the total of Northwest.

Q2. What is the relationship between annual leave taken and bonus?

Related Entity & Relationships

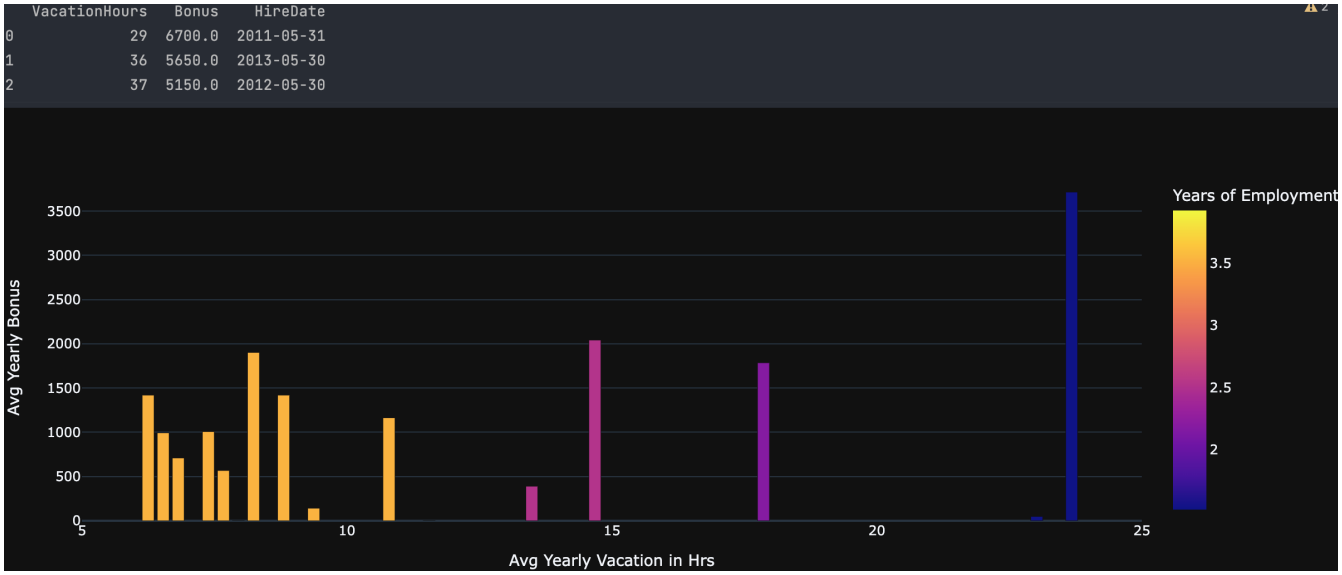

SalesPerson /* Sales representative current information. */
TerritoryID /* Territory currently assigned to. Foreign key to SalesTerritor... *
SalesQuota /* Projected yearly sales. */ money
Bonus /* Bonus due if quota is met. */ money
CommissionPct /* Commision percent received per sale. */ smallmoney
SalesYTD /* Sales total year to date. */ money
SalesLastYear /* Sales total of previous year. */ money
rowguid /* ROWGUIDCOL number uniquely identifying the record. Used to s
ModifiedDate /* Date and time the record was last updated. */ datetime
BusinessEntityID /* Primary key for SalesPerson records. Foreign key to Er
BusinessEntityID
Employee /* Employee information such as salary, department, and title. */
NationalIDNumber /* Unique national identification number such as a social
LoginID /* Network login. */ nvarchar(256)
OrganizationNode /* Where the employee is located in corporate hierarchy. *
OrganizationLevel /* The depth of the employee in the corporate hierarchy. *
JobTitle /* Work title such as Buyer or Sales Representative. */ nvarchar(50)
BirthDate /* Date of birth. */ date
MaritalStatus /* M = Married, S = Single */ nchar(1)
Gender /* M = Male, F = Female */ nchar(1)
HireDate /* Employee hired on this date. */ date
SalariedFlag /* Job classification. 0 = Hourly, not exempt from collective ba...
VacationHours /* Number of available vacation hours. */ smallint
SickLeaveHours /* Number of available sick leave hours. */ smallint
CurrentFlag /* 0 = Inactive, 1 = Active */ unknown
rowguid /* ROWGUIDCOL number uniquely identifying the record. Used to su...
ModifiedDate /* Date and time the record was last updated. */ datetime
BusinessEntityID /* Primary key for Employee records. Foreign key to Busine

SQL

```
SELECT emp.VacationHours, sp.Bonus, emp.HireDate
FROM HumanResources.Employee AS emp
         INNER JOIN Sales.SalesPerson AS sp ON emp.BusinessEntityID =
sp.BusinessEntityID
ORDER BY Bonus DESC, VacationHours DESC
```

Output & Visualisation

	VacationHours	Bonus	HireDate
0	29	6700.0	2011-05-31
1	36	5650.0	2013-05-30
2	37	5150.0	2012-05-30



Assumptions

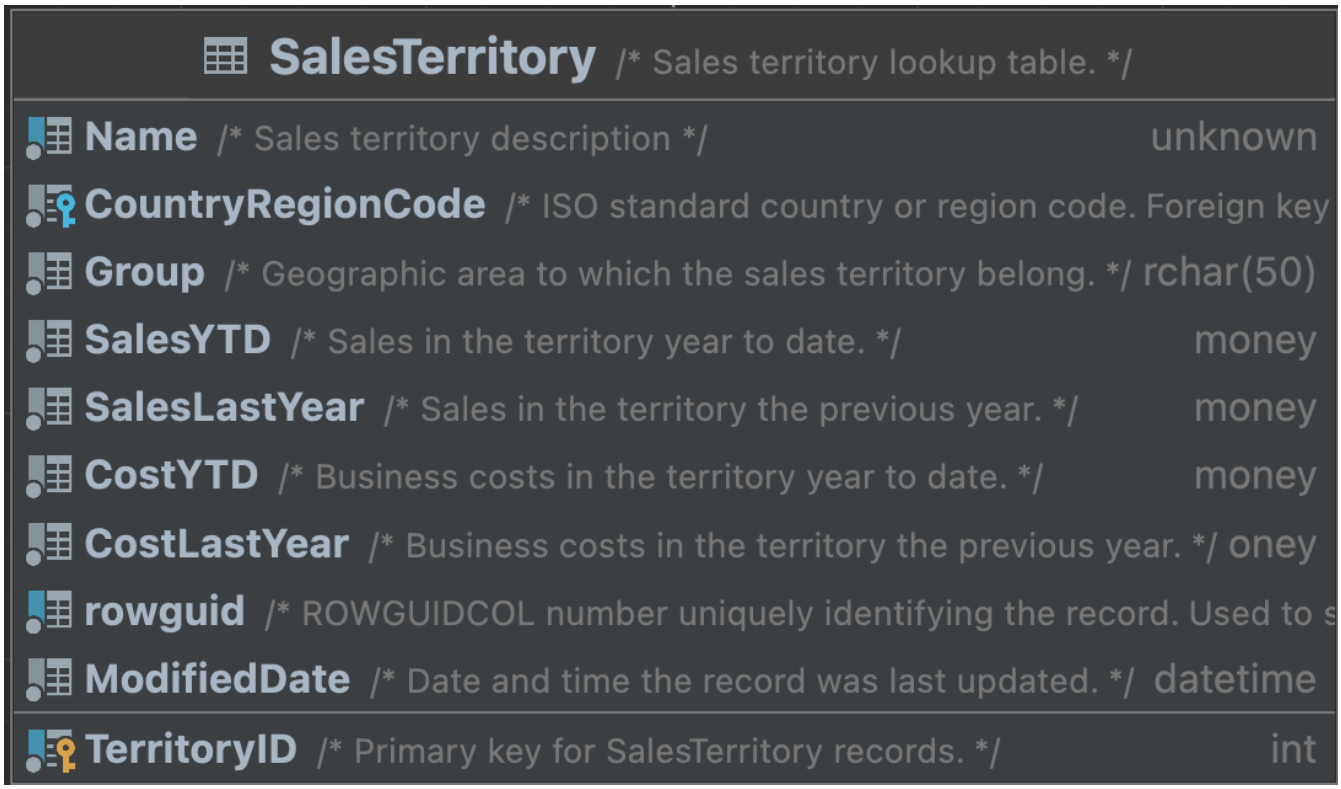
- The date that the visualization was run is 6/12/2014. This was decided due to the database being frozen midway through 2014 and that using the current date would distort the results.
- We assume that employees no longer in the company don't exist in the Employee table as there is no Fired Date values in the database.
- We only compared salespeople as they are the only ones who are paid bonuses.

Answer & Further Insights

- There is no clear pattern of any relationship between average Vacation Time and average Bonus.
- Employees of similar Employment lengths have similar patterns of Vacation time taken and Bonus paid.

Q3. What is the relationship between Country and Revenue?

Related Entity & Relationships



SQL

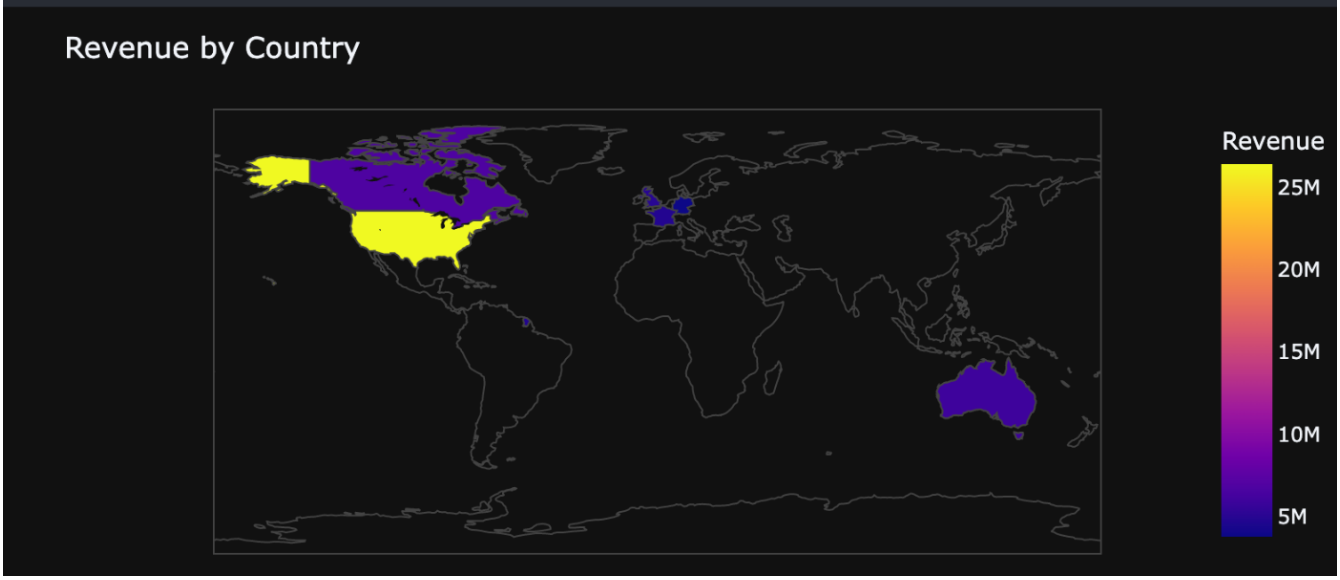
```
SELECT CountryRegionCode AS CountryCode, SUM(SalesYTD) AS Revenue
FROM Sales.SalesTerritory
GROUP BY CountryRegionCode
ORDER BY Revenue DESC
```

Assumptions

- SalesYTD instead of SalesLastYear
- SalesYTD amount is in USD

Output & Visualisation

	CountryCode	Revenue	Country
0	US	2.641106e+07	United States
1	CA	6.771829e+06	Canada
2	AU	5.977815e+06	Australia
3	GB	5.012905e+06	Great Britain
4	FR	4.772398e+06	France
5	DE	3.805202e+06	Germany

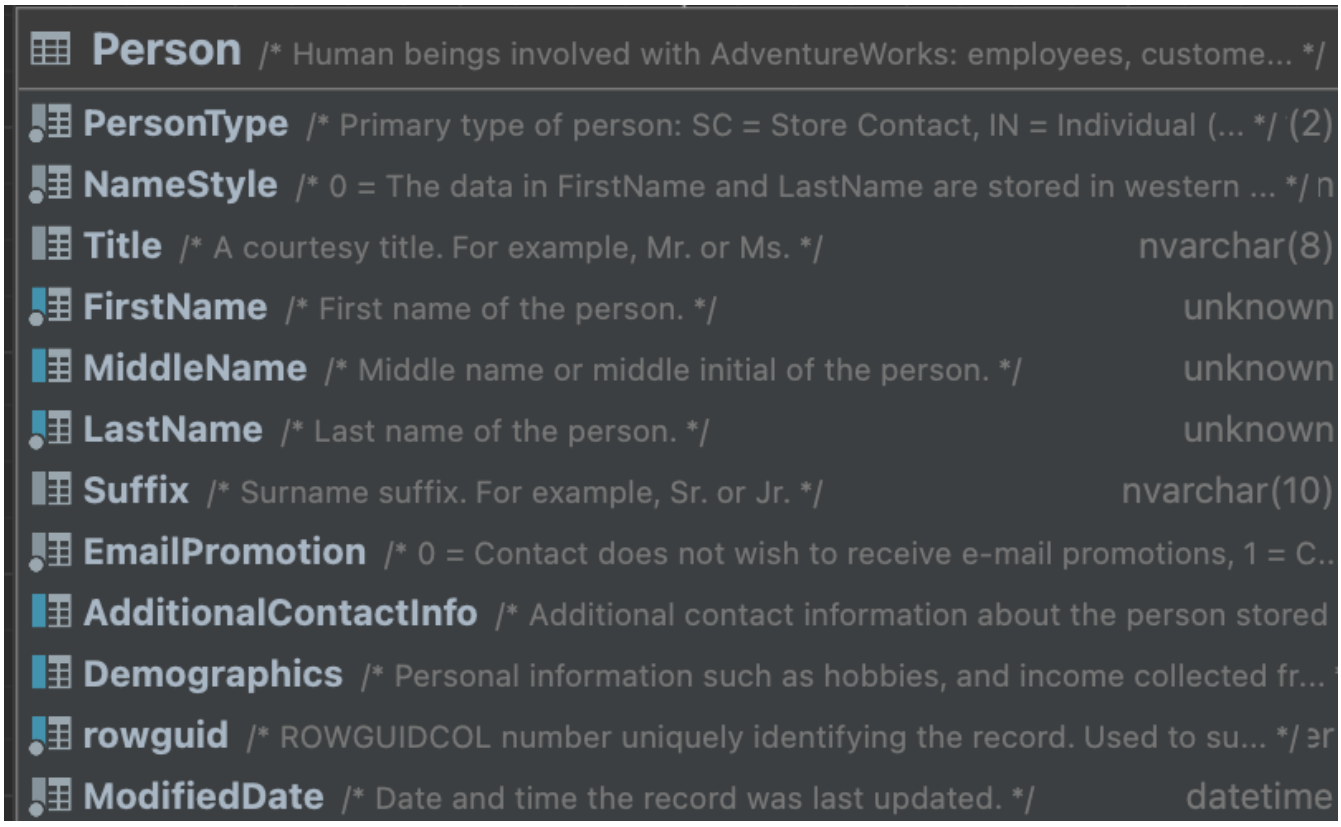


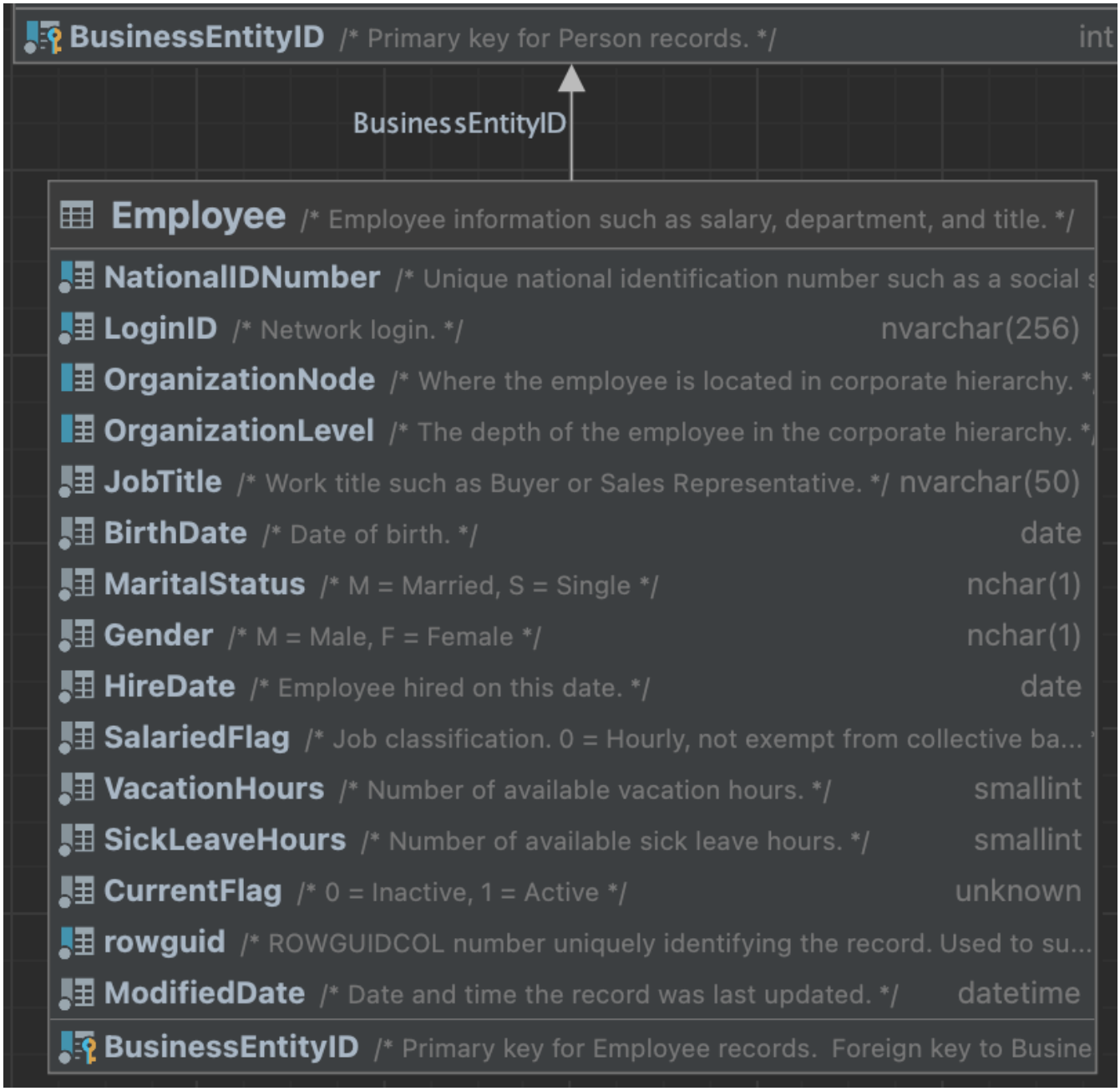
Answer & Further Insights

- United States Canada Australia Great Britain France Germany

Q4. What is the relationship between sick leave and Job Title (PersonType)?

Related Entity & Relationships





SQL

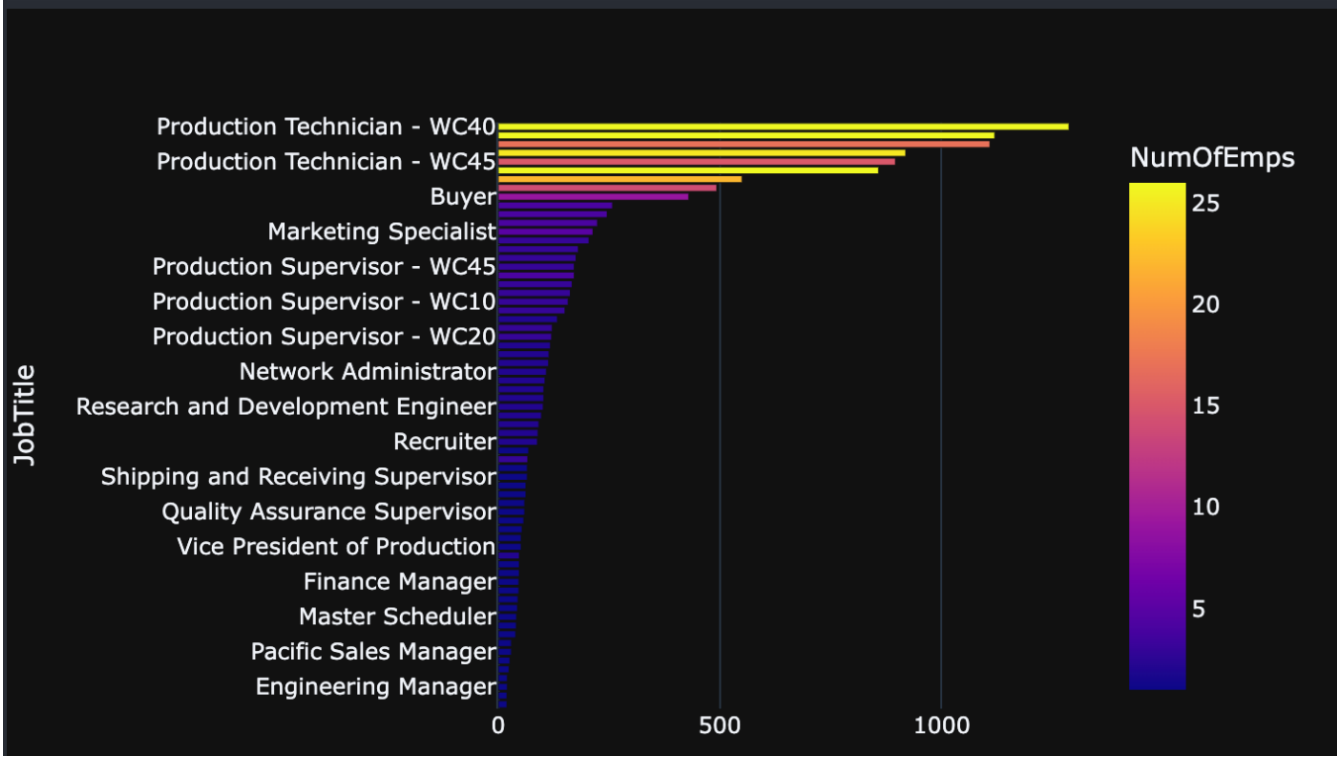
```
SELECT p.PersonType, emp.JobTitle, SUM(emp.SickLeaveHours) AS
TotalSickHours, COUNT(*) AS NumOfEmps
FROM HumanResources.Employee emp
        INNER JOIN Person.Person p ON emp.BusinessEntityID = p.
BusinessEntityID
GROUP BY p.PersonType, emp.JobTitle
ORDER BY TotalSickHours DESC
```

Assumptions

- Count of Employees to accurately represent the TotalSickHours
- All employees are still working with the company.

Output & Visualisation

	PersonType	JobTitle	TotalSickHours	NumOfEmps
0	EM	Production Technician - WC40	1287	26
1	EM	Production Technician - WC50	1120	26
2	EM	Production Technician - WC10	1109	17
3	EM	Production Technician - WC30	919	25
4	EM	Production Technician - WC45	896	15
..	...	...	...	...
62	EM	Vice President of Sales	25	1
63	EM	Senior Design Engineer	21	1
64	EM	Engineering Manager	21	1
65	EM	Chief Financial Officer	20	1
66	EM	Vice President of Engineering	20	1

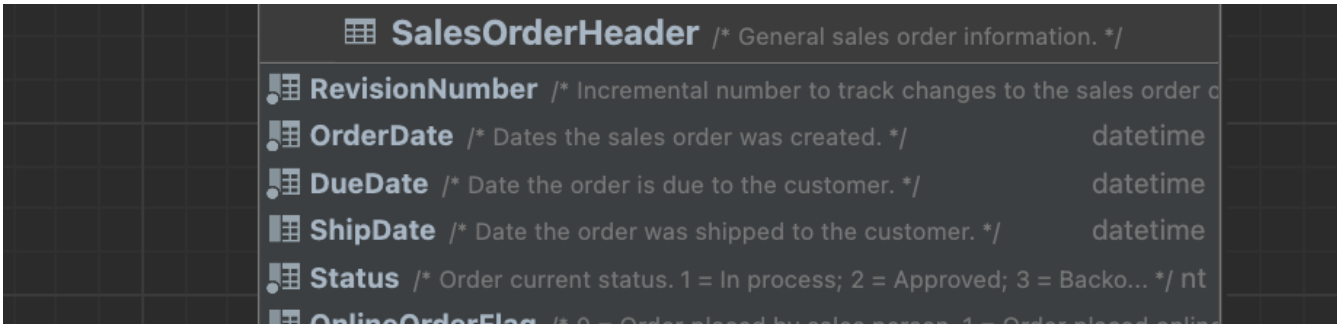


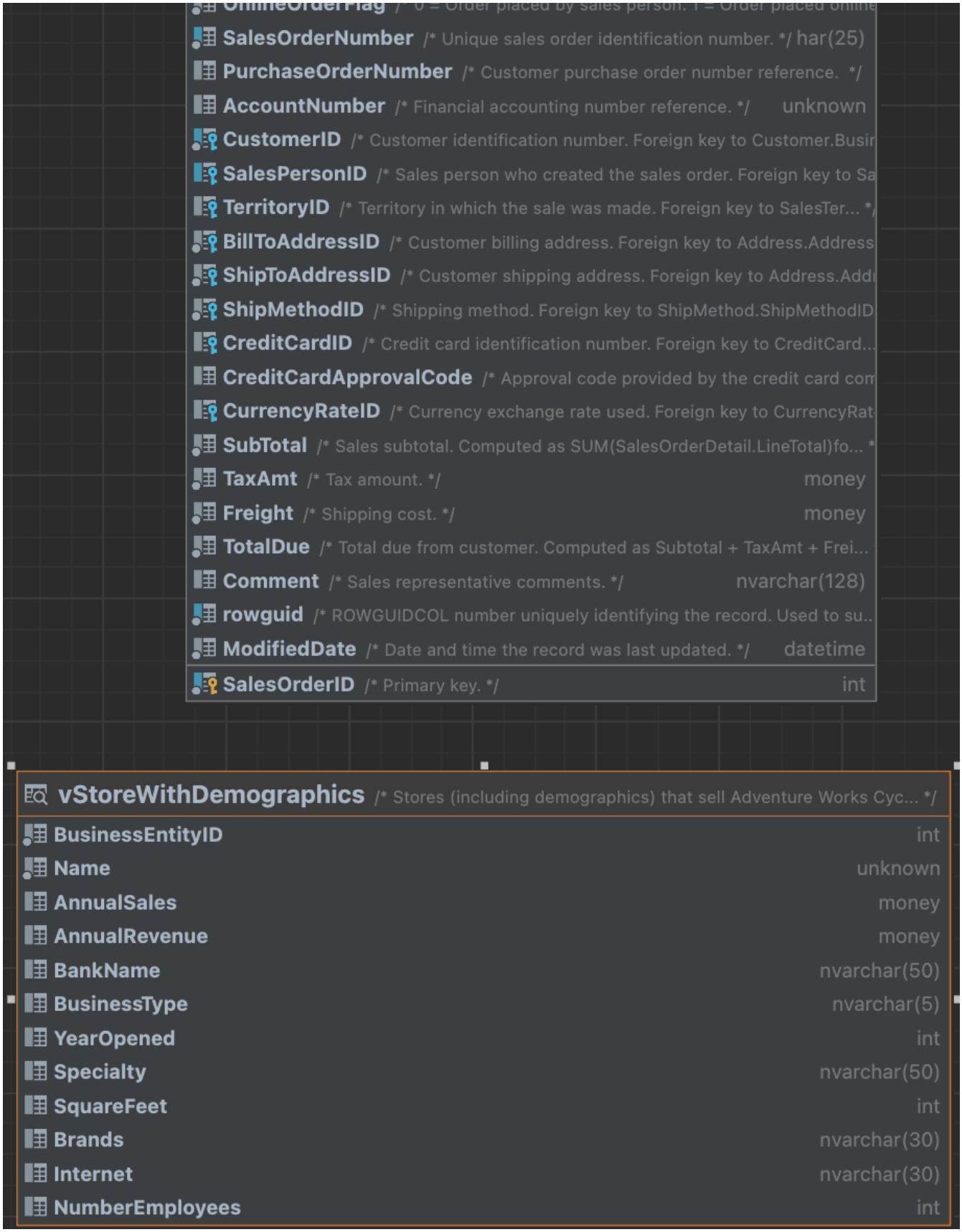
Answer & Further Insights

- Production Technicians WC 40 have taken the most sick leaves in the company - 1287
- Production Technicians WC 50 have taken the second most in the company with 1120 leaves
- Production Technicians WC 10 stand third with 1109 leaves

Q5. What is the relationship between store trading duration and revenue?

Related Entity & Relationships




SalesOrderNumber /* Unique sales order identification number. */ har(25)
PurchaseOrderNumber /* Customer purchase order number reference. */
AccountNumber /* Financial accounting number reference. */ unknown
CustomerID /* Customer identification number. Foreign key to Customer.Busir
SalesPersonID /* Sales person who created the sales order. Foreign key to Sa
TerritoryID /* Territory in which the sale was made. Foreign key to SalesTer... *
BillToAddressID /* Customer billing address. Foreign key to Address.Address
ShipToAddressID /* Customer shipping address. Foreign key to Address.Addı
ShipMethodID /* Shipping method. Foreign key to ShipMethod.ShipMethodID
CreditCardID /* Credit card identification number. Foreign key to CreditCard...
CreditCardApprovalCode /* Approval code provided by the credit card com
CurrencyRateID /* Currency exchange rate used. Foreign key to CurrencyRat
SubTotal /* Sales subtotal. Computed as SUM(SalesOrderDetail.LineTotal)fo... *
TaxAmt /* Tax amount. */ money
Freight /* Shipping cost. */ money
TotalDue /* Total due from customer. Computed as Subtotal + TaxAmt + Frei...
Comment /* Sales representative comments. */ nvarchar(128)
rowguid /* ROWGUIDCOL number uniquely identifying the record. Used to su..
ModifiedDate /* Date and time the record was last updated. */ datetime
SalesOrderID /* Primary key. */ int
vStoreWithDemographics /* Stores (including demographics) that sell Adventure Works Cyc... */
BusinessEntityID int
Name unknown
AnnualSales money
AnnualRevenue money
BankName nvarchar(50)
BusinessType nvarchar(5)
YearOpened int
Specialty nvarchar(50)
SquareFeet int
Brands nvarchar(30)
Internet nvarchar(30)
NumberEmployees int

SQL

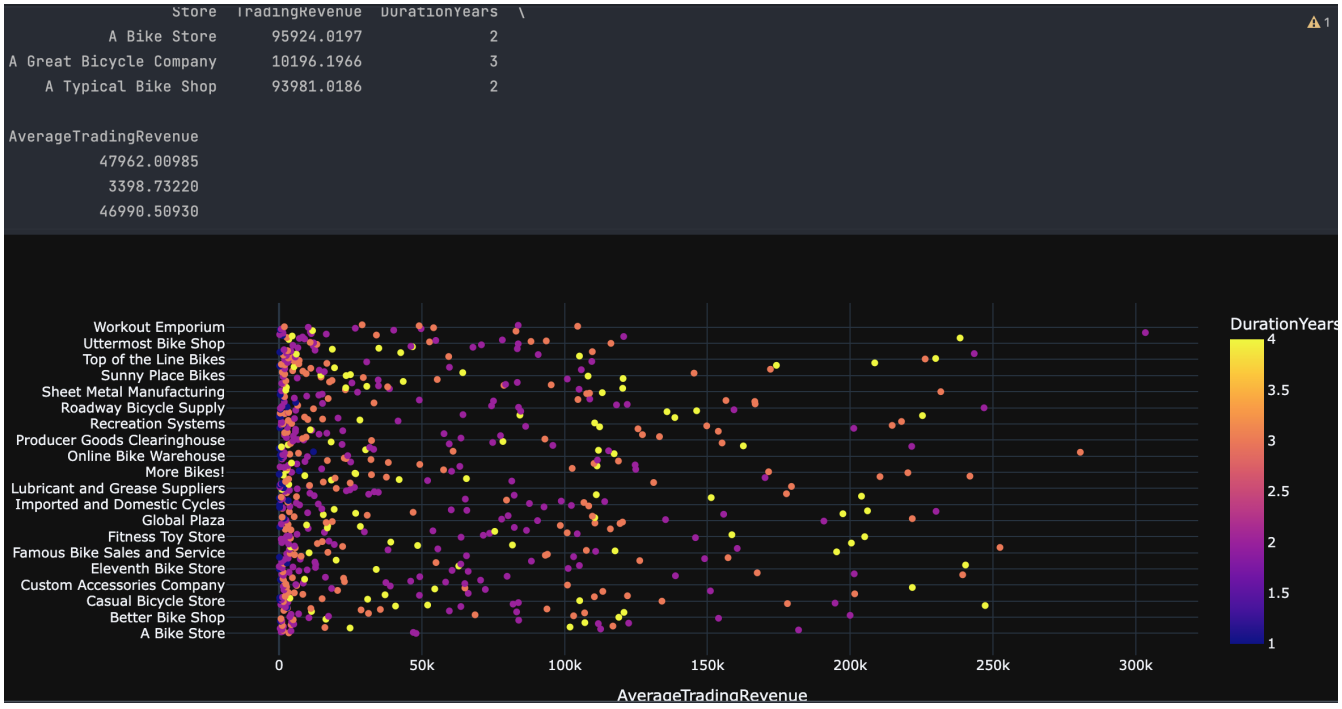
```
WITH YRPS AS (SELECT s.Name AS Store, YEAR(ShipDate) AS Year, SUM
(TotalDue) AS StoreRevenue
              FROM Sales.SalesOrderHeader sod
                       INNER JOIN Sales.Customer c ON sod.CustomerID
= c.CustomerID
                       INNER JOIN Sales.Store s ON c.StoreID = s.
BusinessEntityID
              GROUP BY s.Name, YEAR(ShipDate))

SELECT y.Store, SUM(y.StoreRevenue) AS TradingRevenue, COUNT(*) AS
DurationYears
FROM YRPS y
WHERE EXISTS (SELECT 1
             FROM Sales.vStoreWithDemographics v
             WHERE y.Store = v.Name
               AND y.Year > v.YearOpened)
GROUP BY y.Store
```

Assumptions

- The Store Opening & Closing data is not available.
- We're using number of years for available total revenue data.
- That should give is the relationship between how long the store has been in operation and it's total revenue for that period.

Output & Visualisation

```
                 Store  TradingRevenue  DurationYears  \
          A Bike Store      95924.0197              2
A Great Bicycle Company      10196.1966              3
    A Typical Bike Shop      93981.0186              2

AverageTradingRevenue
          47962.00985
           3398.73220
          46990.50930
```

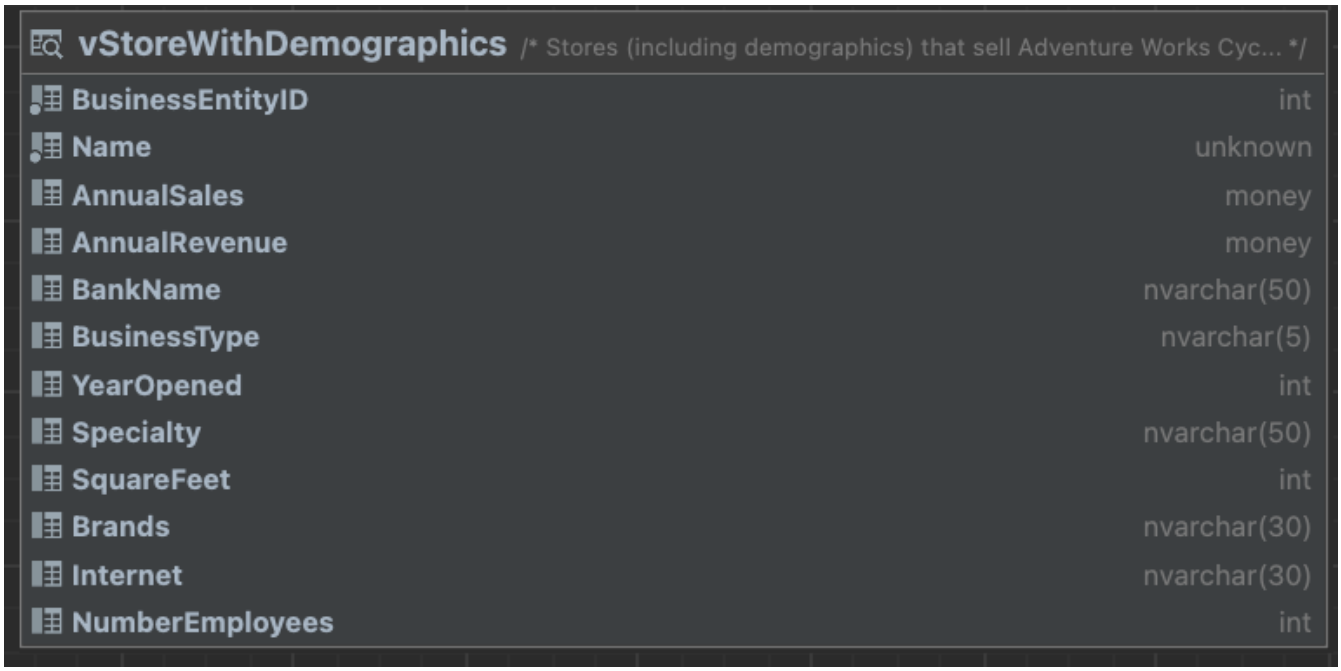


Answer & Further Insights

- Number of years in market presence has no effect or very less impact on revenue generation.
- Other factor consideration.

Q6. What is the relationship between the size of the stores, number of employees and revenue?

Related Entity & Relationships



SQL

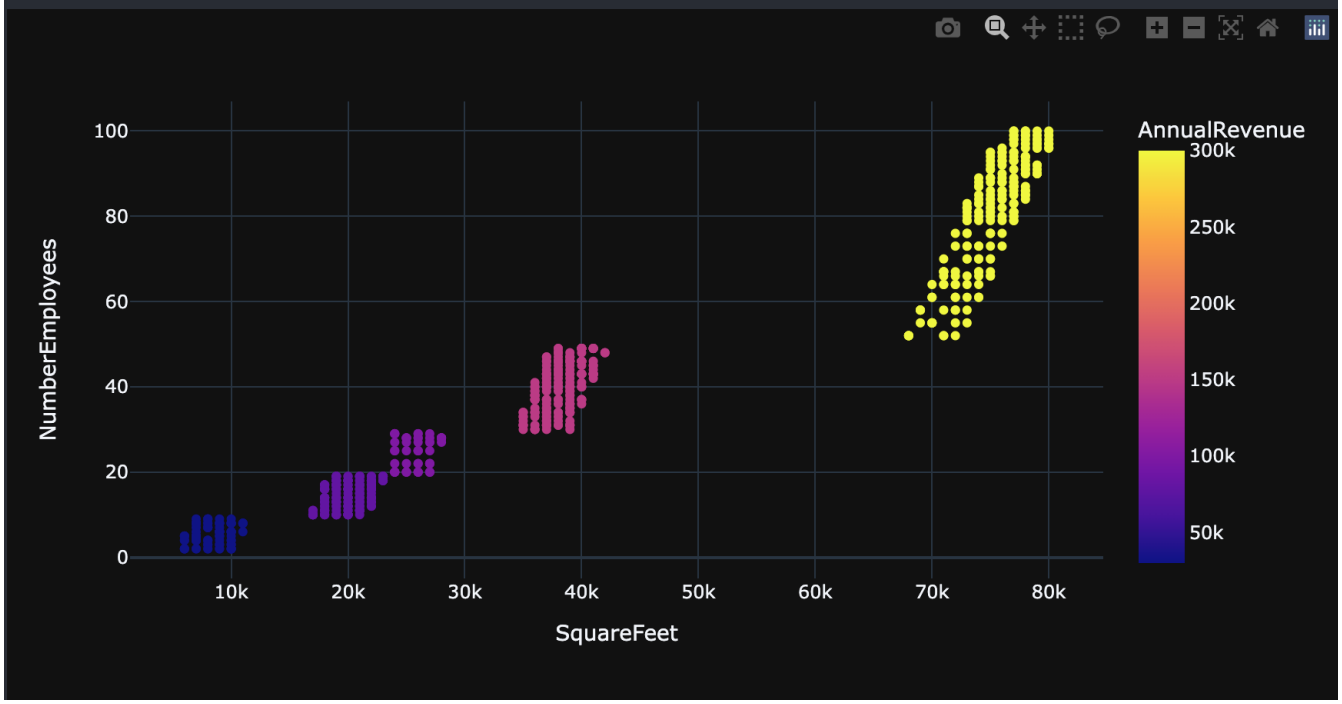
```
SELECT BusinessEntityID, Name, AnnualRevenue, SquareFeet,
NumberEmployees
FROM Sales.vStoreWithDemographics
```

Assumptions

- we have data from view vStoreWithDemographics which have data from xmls link of sales.store

Output & Visualisation

```
   BusinessEntityID                          Name  AnnualRevenue  \
0               292           Next-Door Bike Store        80000.0
1               294  Professional Sales and Service        80000.0
2               296                 Riders Company        80000.0

   SquareFeet  NumberEmployees
0       21000               13
1       18000               14
2       21000               15
```



Answer & Further Insights

- No linear relationship
- Few store revenue low with higher number of employees.
- Number of employee and size of store has positive impact on revenue.