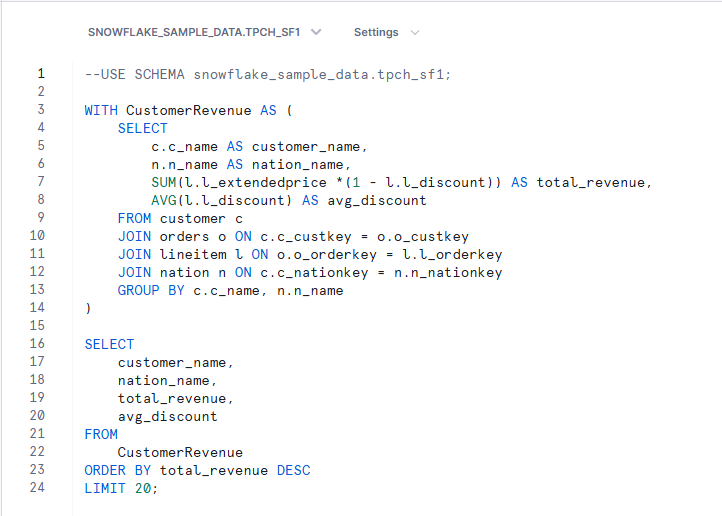
# **SQL queries**

Pro všechny 3 queries jsem zvolil CTE, aby to bylo přehlednější a čitelnější.

## 1, Customer Revenue



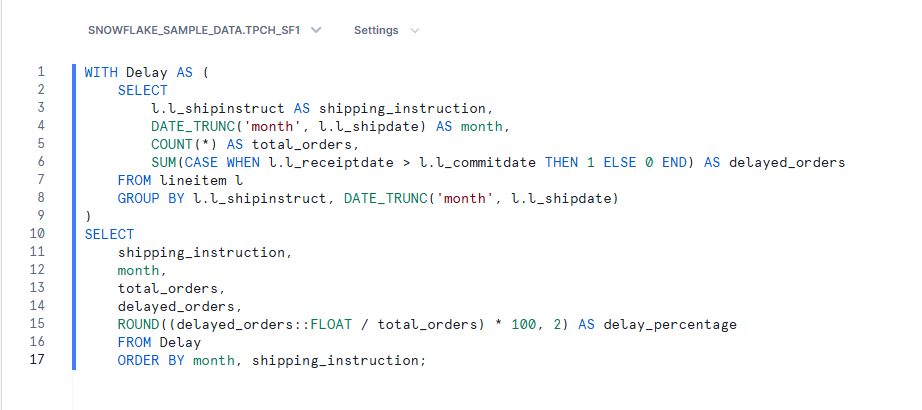
Selektuji customer\_name, nation\_name, total\_revenue a avg\_discount.

Čistý výnos(total\_revenue) počítám dle vzorce: čistý výnos = extendedprice \* ( 1 - discount).

Následně v kódu spojuji customer, orders, lineitem a nation tabulky a seskupuji data dle customer\_name a nation\_name.

V následné části kódu selectuji všechny sloupce z mezivýsledku, seřazuji dle total\_revenue v sestupném pořadí( od nejvyššího po nejnižší) a limituji to na prvních 20 výsledků.

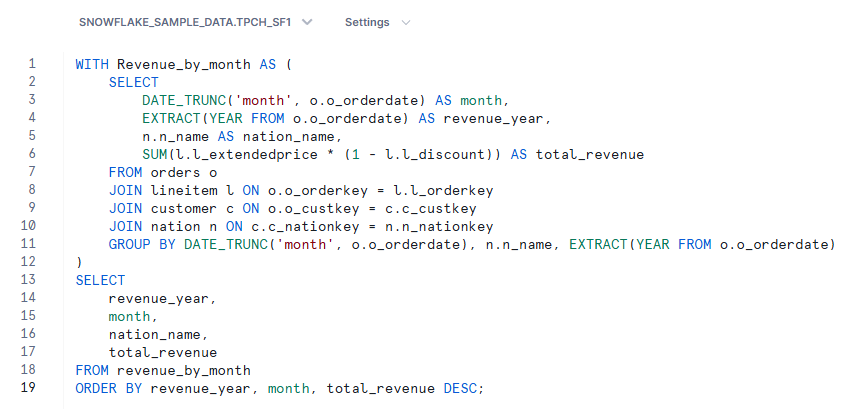
## 2, Delivery delays



Selektuji shipping\_instruction, sjednocuji shipdate napříjkad z 1996-08-22 na 1996-08-01 kvůli GROUP BY a dalšímu zjednodušení při počítání delay\_percentage, celkový počet objednávek pro kombinaci shipping\_instruction a month. Pak výpočet počtu zpožděných objednávek z logiky, pokud je datum obdržení zásilky(receiptdate) větší než datum slíbeného dovozu(commitdate), tak přičtu 1 do delayed orders.

V hlavním dotaze ještě počítám procento zpožděných objednávek a ukládám pod delay\_percentage, zaokrouhleno na 2 desetinná místa.

## 3, Revenue



Sednocuji orderdate napříjkad z 1996-08-22 na 1996-08-01, extrahuji rok z datumu pro případnou jednodušší vizualizaci, nation\_name, čistý výnos(total\_revenue) počítám dle vzorce: čistý výnos = extendedprice \* ( 1 - discount). Seskupuji data dle DATE\_TRUNC('month', o.o\_orderdate), n\_name a EXTRACT(YEAR FROM o.o\_orderdate). Následně seřazuji výsledky dle revenue\_year, month a total\_revenue sestupně.

# **Summarizing the insights**

## Customer revenue

### Average Discount

The average discount applied across TOP10 customers is **5.10%**, with the highest being **5.44%** and the lowest **4.74%.** The small range indicates relatively consistent discount policies among TOP10 customers.

### Total revenue by TOP 10 customers

The TOP10 customers contribute a combined revenue of **61.48 million** with an average revenue of **6.15 million** per customer.

### Top customers

Customer#000143500 is the leading contributor, generating 6.76 million in revenue.

Customers #00095257, #00097115, and #00131113 follow closely, each contributing between 6.29 and 6.08 million.

### Discount for TOP3 customers.

The average applied discount is 5.10%. If we take a look at TOP3 contributors, the first one has a 5.21% discount, the second one 5.06% and third 5.31%. Customer with the highest (5.44%) discount is at 6th position with 6.06 million in revenue for example.

## 

## 

## 

## Delivery delays

### KPI

### Number of orders

There are **6.0 million total orders** in the dataset.

### Delayed orders

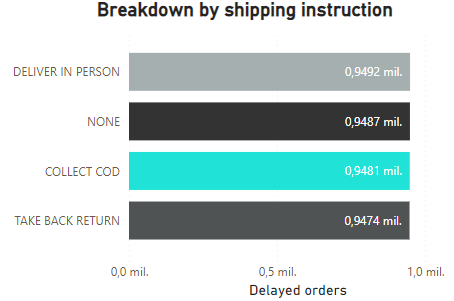
Almost **3.8 million** **of the orders are delayed**, which is a significant number according to total orders.

### Average delayed percentage

The average percentage of delayed orders is **63.36%!**

### Breakdown by shipping instruction

COLLECT COD, DELIVER IN PERSON, NONE, and TAKE BACK RETURN all have approximately 0.95 million delayed orders. Delayed orders are not strongly dependent on the shipping method. **So other factors might be contributing to delays more than the shipping instructions.**

****

### 

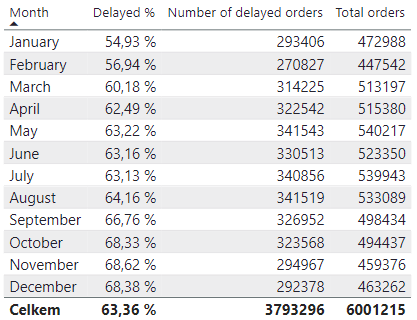
### 

### Average delayed percentage by month

Monthly delayed orders(AVG) ranges **from 0.55% in January to 0.69%** in November. The delays increase slightly towards the end of the year, especially in November (0.69%) and December (0.68%), indicating possible seasonal obstacles.

### Number of delayed orders vs total orders

Months like July and August shows higher delayed orders despite having relatively consistent total orders.



### Percentage of delayed orders

The percentage of delayed orders started low in 1992 and settled on around **63%** from 1993 till 1997. **A significant increase is visible in 1998**, suggesting a potential operational or external disruption in that year.

## Total Revenue

### Revenues by country and year

#### Highest revenue in a single year

* **1996, France:** 1,388,138,667. - France has the highest revenue in multiple years, making it a key contributor.
* **1996, Jordan:** 1,378,613,650.
* **1994, Russia:** 1,377,606,125.
* **1993, Canada:** 1,374,466,649.

#### Highest monthly revenue

* **1996, December, France:** 128,769,431. - This result indicates that December( also July) is a key month for France.
* **1993, January, Mozambique:** 126,515,605.
* **1992, October, Jordan:** 125,840,189.
* **1995, August, Mozambique:** 125,680,939.

#### Highest sum of revenues in specific month

* **July, France:** 821,397,524.
* **May, Russia:** 820,298,481.
* **January, Mozambique:** 817,448,048.
* **March, Canada:** 814,208,889.

#### Total sum of revenues

* **France:** 8,960,205,392.
* **Indonesia:** 8,945,275,218.
* **Russia:** 8,925,318,302.
* **Mozambique:** 8,892,984,086.

#### Lowest revenue in a single year

* **1998, Kenya:** 751,629,473.
* **1998, Argentina:** 753,325,831.
* **1998, Saudi Arabia:** 756,114,953.
* **1998, Japan:** 761,171,763.

#### Lowest monthly revenue

* **1998, August, Russia:** 4,967,293.
* **1998, August, Saudi Arabia:** 5,608,101.
* **1998, August, China:** 5,657,904.
* **1998, August, Argentina:** 5,677,581.

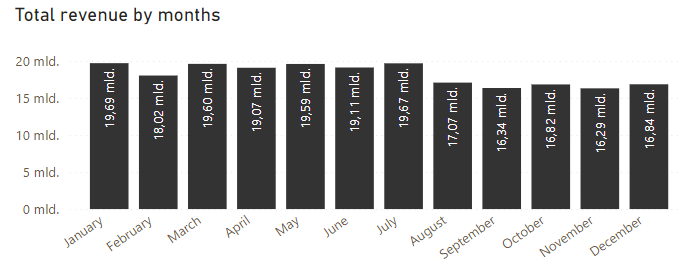
#### Lowest sum of revenues in specific month

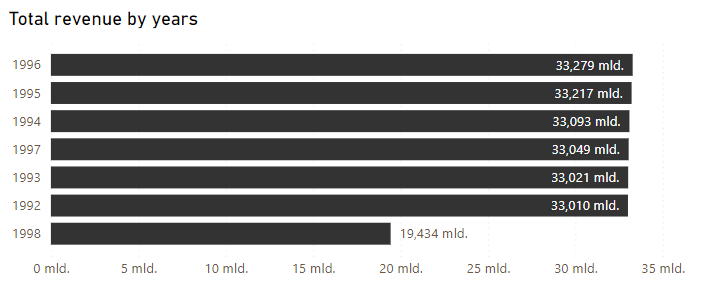
* **November, Egypt:** 622,609,121.
* **September, United Kingdom:** 623,121,227.
* **November, Iran:** 624,784,525.
* **November, Peru:** 627,757,854.

#### Lowest sum of revenues

* **Egypt:** 8,582,234,483.
* **Kenya:** 8,554,593,974.
* **Iraq:** 8,550,329,961.
* **Saudi Arabia:** 8,472,676,397.

### Total revenue by months



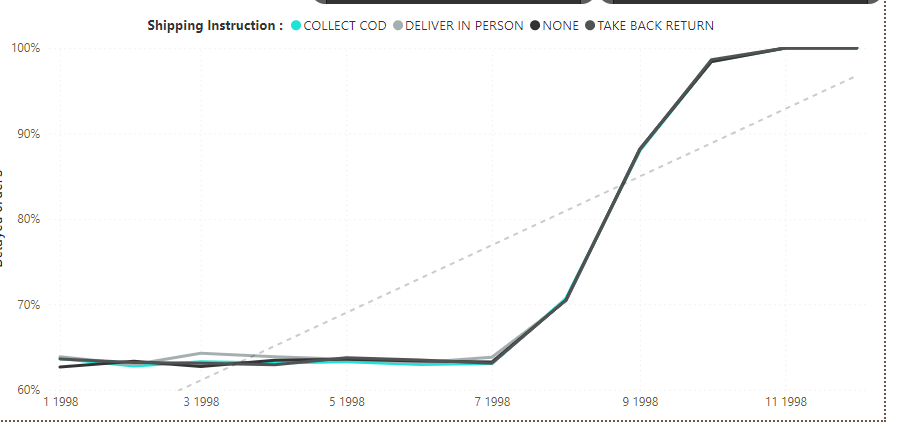
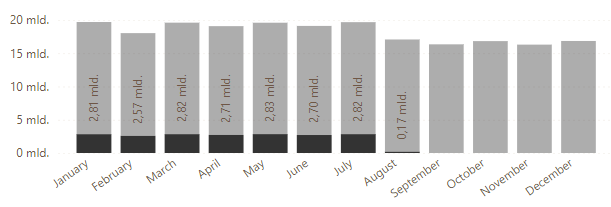


#### Highest revenue months

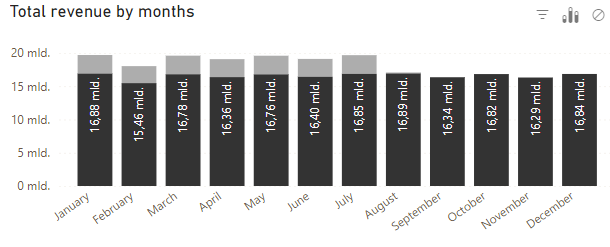
* **January achieved the highest** total revenue of 19.69 billion followed closely by July with 19.67 billion.
* These months represent peak periods. Possibly caused by holiday related activity.

#### Lowest revenue months

Lowest revenue months are October, November and December. These months have the lowest total revenues. It's caused by data which ends at 1998 August, probably the end of business due to delayed deliveries increasing from ~63% in July to 100% in November 1998.



### Seasonal Trends

Revenues peak in January and July. Indicating key periods. If we ignore the year 1998, August is also in the trends. On the other hand, February is the weakest by far. Notable thing is that the performance over the year is very consistent.

### Regional Contributions:

Europe: France and Russia consistently dominate revenues, especially at the end of the year.

Africa: Mozambique shows strong revenues in January and August, Egypt in July.

Asia: Indonesia is a stable contributor, especially during summer months. Iran is pretty strong in december.

America: Canada and Brazil achieve high revenues in spring and at the end of the year.

# AI-Driven Improvements

Predicting and mitigating shipping delays.

Delivery is delayed in 63.36% of all orders and slightly vary by delivery type and by time period. These delays affect customer satisfaction and probably increase logistical costs.

So the proposed AI Solution is : An AI driven predictive model that could analyze historical data to forecast the probability of future delayed deliveries.The key factors for analysis could be Month, Shipping\_instruction, delay\_perc and Delayed orders. It could be used to identify high-risk time periods and delivery types.

Mains benefits of this AI solution would be:

1. Proactive issue solution - teams can expect delays and adjust working resources
2. Efficiency - AI can recommend some logistic optimizations such as reallocating resources to reduce delays in high risk periods
3. Customer satisfaction (better communication) - informing customers about delays

General implementation would look like

1. Build a machine learning model by using historical data.
2. Visualize predictions and recommendations for decisions
3. Continuously update the model with newest available data for the improved model.