Sales Data Analysis

This project is based on Sales data of an organisation with stores in different countries across the world. The aim of this project is to report the profitability of businesses in different countries in a defined period from 1993 to 2021. The report must answer the following questions:

1. What are the top 3 years in terms of profitability?
2. What are the top 10 countries in terms of profitability?
3. In which month during a year in average is the cost of doing business the lowest?
4. Which day in a week do we generate more sales in average?
5. In which time of day in average do we generate more sales, if we define time ranges as 00:00–05:59, 06:00–11:59, 12:00–17:59, and 18:00–23:59
6. How many transactions could not be matched to a specific country where the sale was made? Can you identify exact transactions to be able to fix them?
7. How many transactions are duplicates, if a combination of date, time, and city is considered a unique identifier of a transaction?

I used two data sets as shown below and is available for use to conduct the analysis via link https://drive.google.com/drive/folders/1zBeoPZGFpMmZefuDpc809vIyfiRD0Ram?usp=drive_link :

1. Excel with 100,000 transactions recorded in a period from 1993–2021.

	A	B	C	D	E	F
1	IDTransaction	Date	Time	City	SalesAmount	CostAmount
2	1	1993/01/01	07:06:18	Spresiano	AED 23 960.00	AED 24 679.00
3	2	1993/01/01	07:29:20	Costa de la Calma	AED 345 595.00	AED 345 595.00
4	3	1993/01/01	14:30:30	Hillsboro	AED 135 434.00	AED 128 662.00
5	4	1993/01/01	16:17:08	Stegna	AED 169 110.00	AED 148 817.00
6	5	1993/01/01	22:53:27	Saint-Jean-du-Cardonnay	AED 288 814.00	AED 271 485.00
7	6	1993/01/01	23:55:31	Andocs	AED 479 058.00	AED 483 849.00
8	7	1993/01/02	10:10:18	Africo Nuovo	AED 252 058.00	AED 292 387.00
9	8	1993/01/02	11:55:15	Fontaine-sur-Somme	AED 164 674.00	AED 164 674.00

Transaction Data

1. Excel with all locations, which includes cities and countries, among other data.

	A	B	C	D	E	F	G	H	I	J
1	IDCity	CountryID	StateID	ProvinceID	CityCode	CityName	TimeZone	Latitude	Longitude	
2	1	108	334		`Amir	`Amir	Asia/Jerusalem	33.1787	35.6208	
3	2	108	2969		`Arugot	`Arugot	Asia/Jerusalem	31.7331	34.776	
4	3	108	333		`Aseret	`Aseret	Asia/Jerusalem	31.8251	34.7474	
5	4	108	2969		`Azriqam	`Azriqam	Asia/Jerusalem	31.7511	34.6976	
6	5	108	2969		`En HaShelosha	`En HaShelosha	Asia/Jerusalem	31.3511	34.4016	
7	6	108	333		`Enat	`Enat	Asia/Jerusalem	32.0828	34.9394	
8	7	108	334		`Evron	`Evron	Asia/Jerusalem	32.9956	35.1009	

Location Data

Data Cleaning

The steps in Data Cleaning:

1. I made a copy of the data on one spreadsheet called FactTransactions. Checked if the data is in table format, check titles if they are correct.
2. I used the Power Query Editor to merge transaction data and location data based on City column on a new sheet called MergedCountryAndCityData. I created a column in FactTransaction sheet called Country and populated it using *=XLOOKUP([@City],'Merged Country And City Data'!$F$2:$F$109450,'Merged Country And City Data'!$M$2:$M$109450).*
3. I also checked for incomplete data and found 47 transactions without location data. I used the formula *=COUNTIFS(FactTransactions!$I$2:$I$100001,"#N/A",Transactions[SalesAmount], ">0").*
4. I checked f there were duplicates using conditional formatting then counted the number of duplicates then I deleted the duplicates.
5. I added new column called Profit and calculated it using formula *=[@SalesAmount]-[@CostAmount].*
6. I *used data from the column Date to a*dd a new column called Years using formula *=YEAR([@Date])*, Month_num using *=[@[Month_Num]]*, Day of the week using *=WEEKDAY([@Date]).*I created another sheet with date details called DateTime ,then used it to find the Month name using *=XLOOKUP([@[Month_Num]],DateTime!$A$2#,DateTime!$B$2:$B$13)* and Day of the week using *=XLOOKUP([@[Day of the week]],DateTime!$D$2#,DateTime!$E$2:$E$8). I added* a column called TimeRange that groups time into 4 groups using *=IF(AND(C2>DateTime!$H$2,C2<DateTime!$I$2),DateTime!$G$2,IF(AND(C2>DateTime!$H$3,C2<DateTime!$I$3),DateTime!$G$3,IF(AND(C2>DateTime!$H$4,C2<DateTime!$I$4),DateTime!$G$4,DateTime!$G$5))).* The data is shown below:

Month_Number	Month
1	January
2	February
3	March
4	April
5	May
6	June
7	July
8	August
9	September
10	October
11	November
12	December

Day_Numbe	Day
6	Friday
7	Saturday
1	Sunday
2	Monday
3	Tuesday
4	Wednesday
5	Thursday

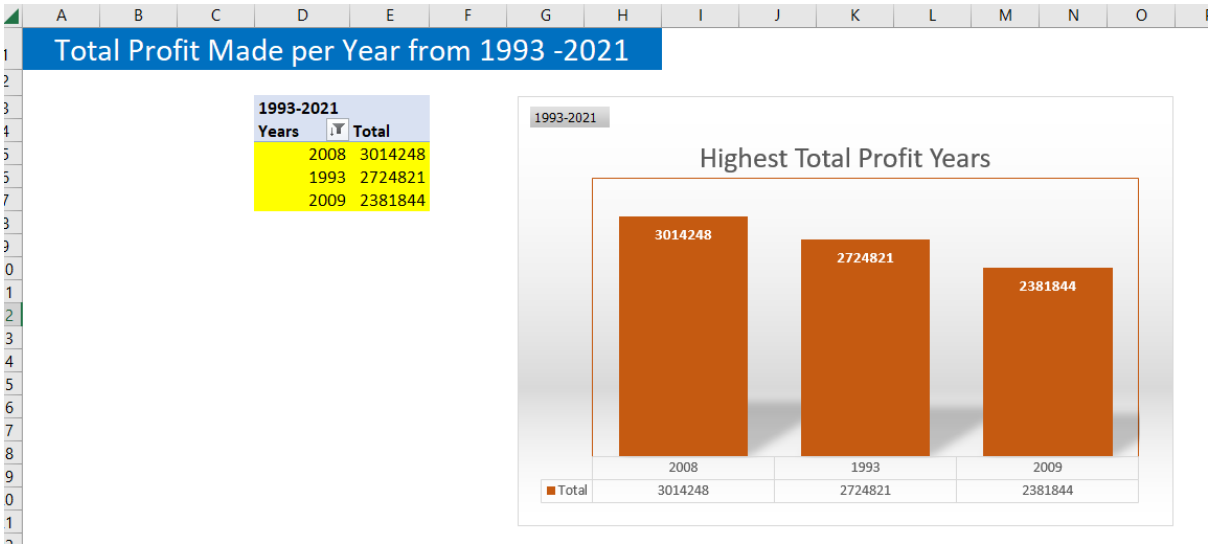
Time	Start	End
After Midnight	00:00	05:59
Morning	06:00	11:59
Afternoon	12:00	17:59
Night	18:00	23:59

Row Labels	Sum of SalesAmount
After Midnight	6301124839
Afternoon	6208559379
Morning	6222910168
Night	6322286193

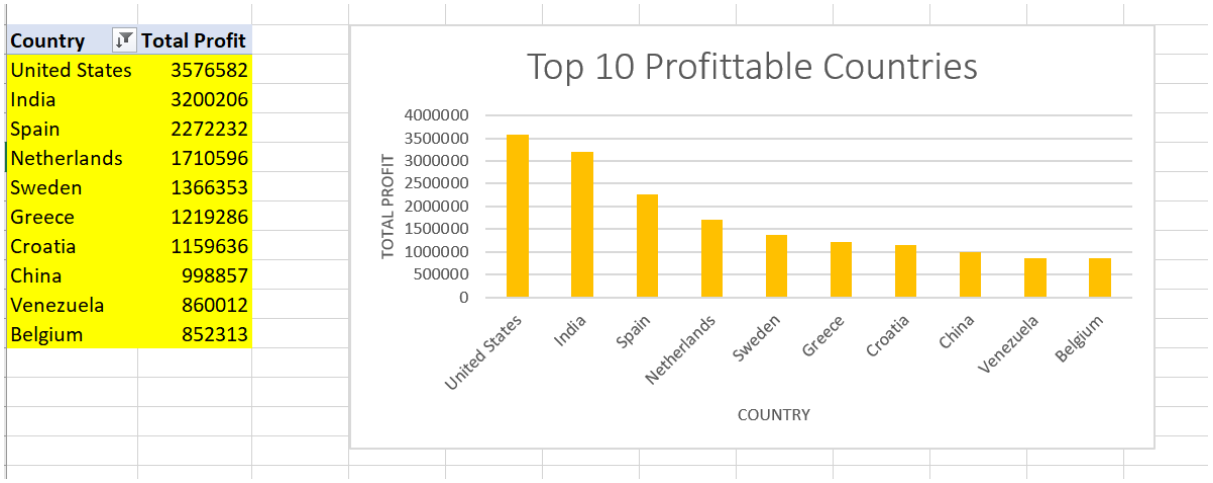
DateTime

Visualize

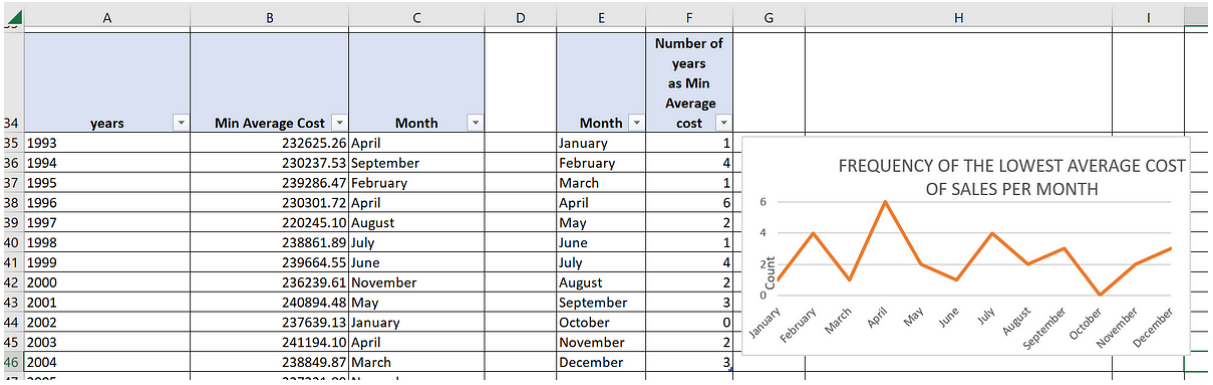
1. I created a pivot table and charts for Top 3 Profitable years as shown below:



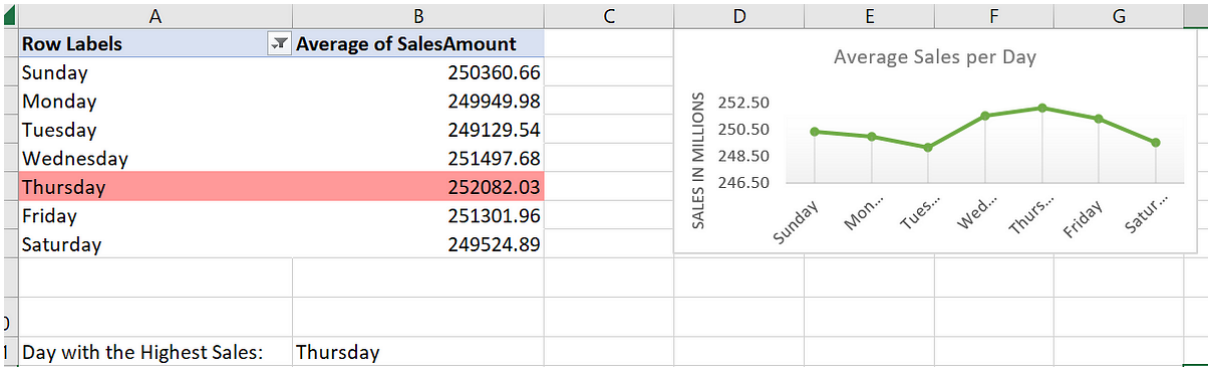
2.I created a pivot table and charts for Top 10 Profitable countries as shown below:



3. For average cost per month each year, i used the pivot table to calculate the least Average cost per year using *=MIN(B3:M3)* and the month using *=OFFSET(INDEX(B3:M3,MATCH(B35,B3:M3,0)),-1,0),* then created another table to calculate the number of times a month has the least average cost using *=COUNTIF($C$35:C63,E35).*

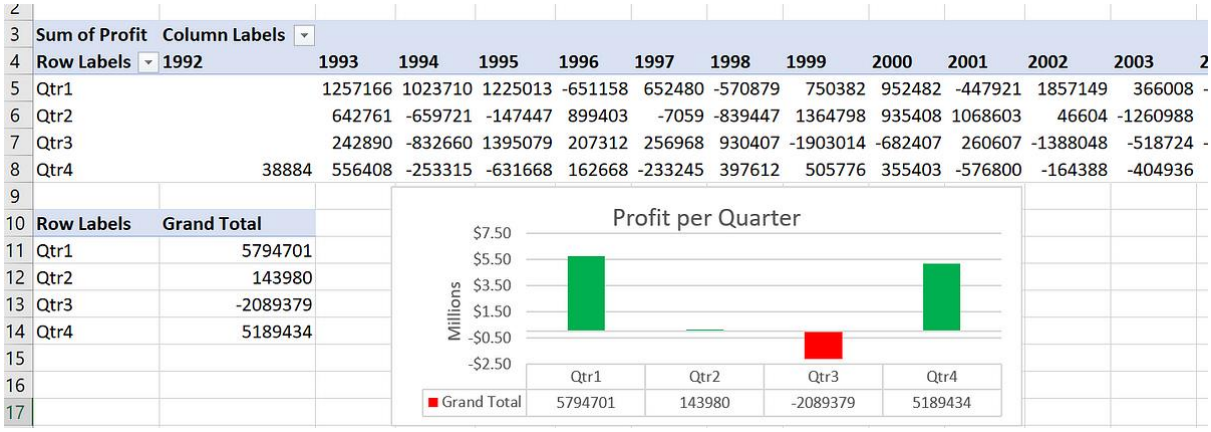


4. I created pivot table and line chart to show average sales per day:



Row Labels	Average of SalesAmount
Sunday	250360.66
Monday	249949.98
Tuesday	249129.54
Wednesday	251497.68
Thursday	252082.03
Friday	251301.96
Saturday	249524.89
Day with the Highest Sales:	Thursday

5. I created a pivot table that shows profit the business makes every year per quarter. Then used the table to create a bar chart that shows total profit/loss per quarter:

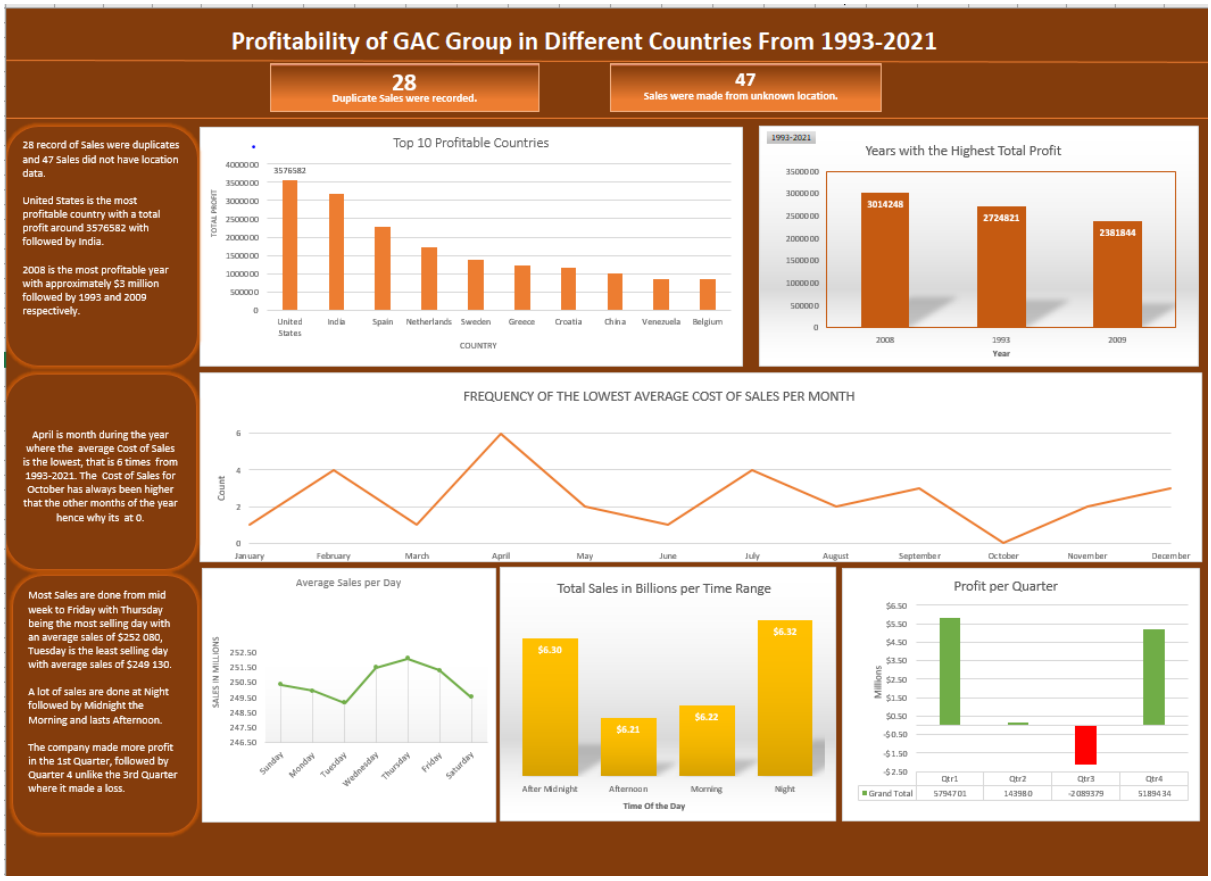


Sum of Profit	Column Labels												
Row Labels	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2
Qtr1		1257166	1023710	1225013	-651158	652480	-570879	750382	952482	-447921	1857149	366008	-
Qtr2		642761	-659721	-147447	899403	-7059	-839447	1364798	935408	1068603	46604	-1260988	
Qtr3		242890	-832660	1395079	207312	256968	930407	-1903014	-682407	260607	-1388048	-518724	-
Qtr4	38884	556408	-253315	-631668	162668	-233245	397612	505776	355403	-576800	-164388	-404936	

Row Labels	Grand Total
Qtr1	5794701
Qtr2	143980
Qtr3	-2089379
Qtr4	5189434

Dashboarding & Reporting

Below is the dashboard I created that shows the Top 10 Profitable countries, Top 3 Profitable years which is 1993,2008 and 2009. The profitable quarter is Q1 followed by Q4, the day where most sales are made which is Thursday.


Profitability of GAC Group in Different Countries From 1993-2021
28
Duplicate Sales were recorded.
47
Sales were made from unknown location.
28 record of Sales were duplicates and 47 Sales did not have location data.
United States is the most profitable country with a total profit around 3576582 with followed by India.
2008 is the most profitable year with approximately $3 million followed by 1993 and 2009 respectively.
Top 10 Profitable Countries
3576582
TOTAL PROFIT
United States
India
Spain
Netherlands
Sweden
Greece
Croatia
China
Venezuela
Belgium
COUNTRY
1993-2021
Years with the Highest Total Profit
3014248
2724821
2381844
2008
1993
2009
Year
April is month during the year where the average Cost of Sales is the lowest, that is 6 times from 1993-2021. The Cost of Sales for October has always been higher that the other months of the year hence why its at 0.
FREQUENCY OF THE LOWEST AVERAGE COST OF SALES PER MONTH
Count
January
February
March
April
May
June
July
August
September
October
November
December
Most Sales are done from mid week to Friday with Thursday being the most selling day with an average sales of $252 080, Tuesday is the least selling day with average sales of $249 130.
A lot of sales are done at Night followed by Midnight the Morning and lasts Afternoon.
The company made more profit in the 1st Quarter, followed by Quarter 4 unlike the 3rd Quarter where it made a loss.
Average Sales per Day
SALES IN MILLIONS
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Total Sales in Billions per Time Range
$6.30
$6.21
$6.22
$6.32
After Midnight
Afternoon
Morning
Night
Time Of the Day
Profit per Quarter
Millions
Qtr1
Qtr2
Qtr3
Qtr4
Grand Total
5794701
143980
-2089379
5189434