In [223…
```
greetings = "Assalam-O-Alaikum!"
print(greetings)
```

```
Assalam-O-Alaikum!
```

Import Libraries

In [224…
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import calendar
```

Import Dataset

In [225…
```
df = pd.read_csv("car_sales_data.csv")
df.head(5)
```

Out[225]:

	Date	Salesperson	Customer Name	Car Make	Car Model	Car Year	Sale Price	Commission Rate	Commission Earned
0	2022-08-01	Monica Moore MD	Mary Butler	Nissan	Altima	2018	15983	0.070495	1126.73
1	2023-03-15	Roberto Rose	Richard Pierce	Nissan	F-150	2016	38474	0.134439	5172.40
2	2023-04-29	Ashley Ramos	Sandra Moore	Ford	Civic	2016	33340	0.114536	3818.63
3	2022-09-04	Patrick Harris	Johnny Scott	Ford	Altima	2013	41937	0.092191	3866.20
4	2022-06-16	Eric Lopez	Vanessa Jones	Honda	Silverado	2022	20256	0.113490	2298.85

In [226…
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2500000 entries, 0 to 2499999
Data columns (total 9 columns):
 #   Column             Dtype  
---  ------             -----  
 0   Date               object 
 1   Salesperson        object 
 2   Customer Name      object 
 3   Car Make           object 
 4   Car Model          object 
 5   Car Year           int64  
 6   Sale Price         int64  
 7   Commission Rate    float64
 8   Commission Earned  float64
dtypes: float64(2), int64(2), object(5)
memory usage: 171.7+ MB
```

In [227…
```
df.describe().astype(float).round(3)
```

Out[227]:

	Car Year	Sale Price	Commission Rate	Commission Earned
count	2500000.000	2500000.000	2500000.000	2500000.000
mean	2015.996	30012.179	0.100	3001.005
std	3.739	11545.137	0.029	1481.467
min	2010.000	10000.000	0.050	501.340
25%	2013.000	20019.000	0.075	1821.710
50%	2016.000	30006.000	0.100	2741.910
75%	2019.000	40022.000	0.125	3978.142
max	2022.000	50000.000	0.150	7494.530

1. What is the total number of records in the dataset?

In [228…
```
records =len(df)
print("Total Number of Records in Dataset = " +str(records))
```

```
Total Number of Records in Dataset = 2500000
```

2. Which car appears most frequently in the dataset?

In [229…
```
most_frequent =df["Car Model"].value_counts().to_frame().reset_index()
most_frequent.columns = ["Car Models", "Total Values"]
most_frequent = most_frequent.sort_values("Total Values", ascending =False)
most_frequent
```

Out[229]:

	Car Models	Total Values
0	Silverado	500842
1	Civic	500498
2	Corolla	500162
3	F-150	499679
4	Altima	498819

In [230…
```
most_frequent_car =most_frequent.iloc[0]
print("Car Model", "'"+ most_frequent_car["Car Models"]+"'", "has most frequently car in dataset = " + str(most
```

```
Car Model 'Silverado' has most frequently car in dataset = 500842
```

3. What is the average sale price of all the cars in the dataset?

In [231…
```
average_sale_price = df["Sale Price"].agg("mean").round(2)
print("Average Sale Price Of All cars In Dataset = $" + str(average_sale_price))
```

```
Average Sale Price Of All cars In Dataset = $30012.18
```

4. Who earned the highest commission?

In [232…
```
highest_commission =df.groupby("Salesperson")["Commission Earned"].agg("sum").to_frame().reset_index().sort_val
highest_commission.head(10)
```

Out[232]:

	Salesperson	Commission Earned
343406	Michael Smith	3658467.56
342657	Michael Johnson	2883629.29
123811	David Smith	2456566.33
213701	James Smith	2448405.59
425509	Robert Smith	2275950.52
343648	Michael Williams	2263033.88
228050	Jennifer Smith	2233631.21
244969	John Smith	2227050.87
96072	Christopher Smith	2222190.97
342667	Michael Jones	2083043.66

In [233…
```
Highest_Earned_person = highest_commission.iloc[0]
print("Most Commission Earned By", "'"+Highest_Earned_person["Salesperson"]+"'", "= $" + str(Highest_Earned_per
```

```
Most Commission Earned By 'Michael Smith' = $3658467.56
```

5. What is the highest commission earned?

In [234…
```
HCE =df[df["Commission Earned"] == df["Commission Earned"].max()]
HCE
```

Out[234]:

	Date	Salesperson	Customer Name	Car Make	Car Model	Car Year	Sale Price	Commission Rate	Commission Earned
363070	2023-02-13	Benjamin Rasmussen	Billy Fernandez	Honda	Corolla	2012	49998	0.149897	7494.53

6. How many sales were made in the year 2022?

In [235…
```
df["Date"] = pd.to_datetime(df["Date"])
```

In [236…
```
df["Sale Year"] = df["Date"].dt.year
```

In [237…
```
SMY = len(df[df["Sale Year"] == 2022])
print("Total Sales were made in year 2022 = " +str(SMY))
```

```
Total Sales were made in year 2022 = 1674502
```

7. What is the total commission earned in the year 2023?

In [238…
```
CE = df[df["Sale Year"] == 2023]
commission_Earned_2023 = CE["Commission Earned"].sum().round(2)
print("Total Commission Earned 2023 = $" + str(format(commission_Earned_2023, ",")))
```

```
Total Commission Earned 2023 = $2,477,578,866.69
```

8. Which car model has the highest average sale price?

In [239…
```
HA = df.groupby("Car Model")["Sale Price"].agg("mean").round(2).to_frame().reset_index().sort_values("Sale Pric
```

```
Highest_Sale_Price = HA.iloc[0]
print("Car Model", "'"+Highest_Sale_Price["Car Model"]+"'", "has Highest Sale Price = $" + str(format(Highest_S
```

```
Car Model 'Silverado' has Highest Sale Price = $30,030.18
```

9. Calculate the total sales for each salesperson.

In [240…
```
df.head(1)
```

Out[240]:

	Date	Salesperson	Customer Name	Car Make	Car Model	Car Year	Sale Price	Commission Rate	Commission Earned	Sale Year
0	2022-08-01	Monica Moore MD	Mary Butler	Nissan	Altima	2018	15983	0.070495	1126.73	2022

In [241…
```
total_sales = df.groupby("Salesperson")["Sale Price"].agg("sum").to_frame().reset_index().sort_values("Sale Pri
total_sales["Sale Price"] = total_sales["Sale Price"].apply(lambda x: '{:,}'.format(x))
total_sales
```

Out[241]:

	Salesperson	Sale Price
343406	Michael Smith	36,806,401
342657	Michael Johnson	29,012,949
123811	David Smith	24,208,781
213701	James Smith	23,905,299
343648	Michael Williams	22,816,628
425509	Robert Smith	22,541,917
228050	Jennifer Smith	22,384,383
96072	Christopher Smith	22,342,430
244969	John Smith	21,970,840
342667	Michael Jones	20,768,839

10. Who made the highest number of sales?

In [242…
```
Highest_Sale = total_sales.iloc[0]
print("Salesperson", "'"+Highest_Sale["Salesperson"]+"'", "has made the highest number of Sale = $" +str(Highes
```

```
Salesperson 'Michael Smith' has made the highest number of Sale = $36,806,401
```

11. Calculate the average commission rate across all sales.

In [243…
```
average_rate =df["Commission Rate"].agg("mean").round(3)
print("The average commission rate across all sales = " + str(average_rate))
```

```
The average commission rate across all sales = 0.1
```

12. Which car make has the highest total commission earned?

In [244…
```
total_earn = df.groupby("Car Make")["Commission Earned"].sum().to_frame().reset_index().sort_values("Commission
total_earn["Commission Earned"] = total_earn["Commission Earned"].apply(lambda x: '{:,}'.format(x))
total_earn = total_earn.iloc[0]
print("Car Manufacturer Comapany", "'"+total_earn["Car Make"]+"'", "has earned the highest total commission = $
```

```
Car Manufacturer Comapany 'Honda' has earned the highest total commission = $1,503,709,359.13
```

13. What is the average commission earned per sale?

In [245…
```
commission_earn = df["Commission Earned"].agg("mean").astype(int)
print("The Average Commission Earned Per Sale = $" +str(commission_earn))
```

```
The Average Commission Earned Per Sale = $3001
```

14. Calculate the total commission earned for each year.

In [246…
```
total_commission_by_year =df.groupby("Sale Year")["Commission Earned"].agg("sum").to_frame().reset_index()
total_commission_by_year["Commission Earned"] = total_commission_by_year["Commission Earned"].apply(lambda x: '
total_commission_by_year
```

Out[246]:

	Sale Year	Commission Earned
0	2022	5,024,934,475.47
1	2023	2,477,578,866.69

In [247…
```
df.head(2)
```

Out[247]:

	Date	Salesperson	Customer Name	Car Make	Car Model	Car Year	Sale Price	Commission Rate	Commission Earned	Sale Year
0	2022-08-01	Monica Moore MD	Mary Butler	Nissan	Altima	2018	15983	0.070495	1126.73	2022
1	2023-03-15	Roberto Rose	Richard Pierce	Nissan	F-150	2016	38474	0.134439	5172.40	2023

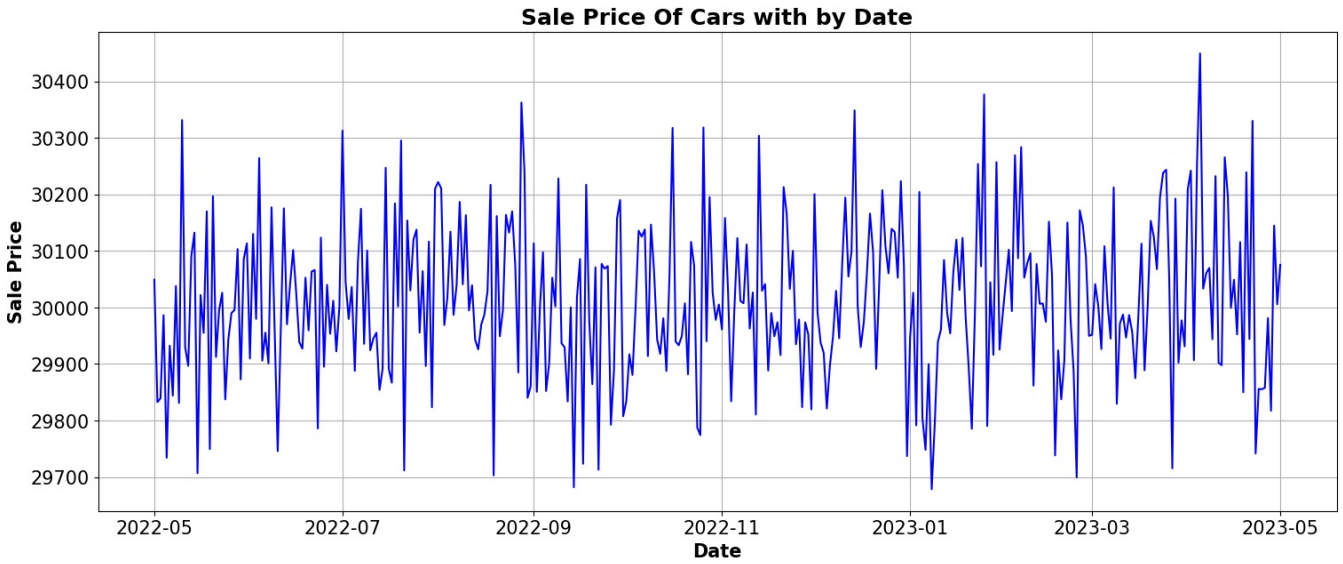
15. Determine the top 3 salespeople based on their total commission earned.

In [248…
```
SPCE =df.groupby("Salesperson")["Commission Earned"].agg("sum").to_frame().reset_index().sort_values("Commissio
SPCE
```

Out[248]:

	Salesperson	Commission Earned
343406	Michael Smith	3658467.56
342657	Michael Johnson	2883629.29
123811	David Smith	2456566.33

In [249…
```
plt.figure(figsize =(18,7))
sns.lineplot(x ="Date", y ="Sale Price", data = df, errorbar =None, color ="blue")
plt.title("Sale Price Of Cars with by Date", weight ="bold", size =18)
plt.xticks(size =15)
plt.yticks(size =15)
plt.xlabel("Date", weight ="bold", size =15)
plt.ylabel("Sale Price", weight ="bold", size =15)
plt.grid()
plt.show()
```



In [250…
```
Sales_by_Date =df.groupby("Date")["Date"].agg("count").to_frame()
Sales_by_Date.columns = ["Car Sold"]
Sales_by_Date= Sales_by_Date.reset_index()
Sales_by_Date.head(10)
```

Out[250]:

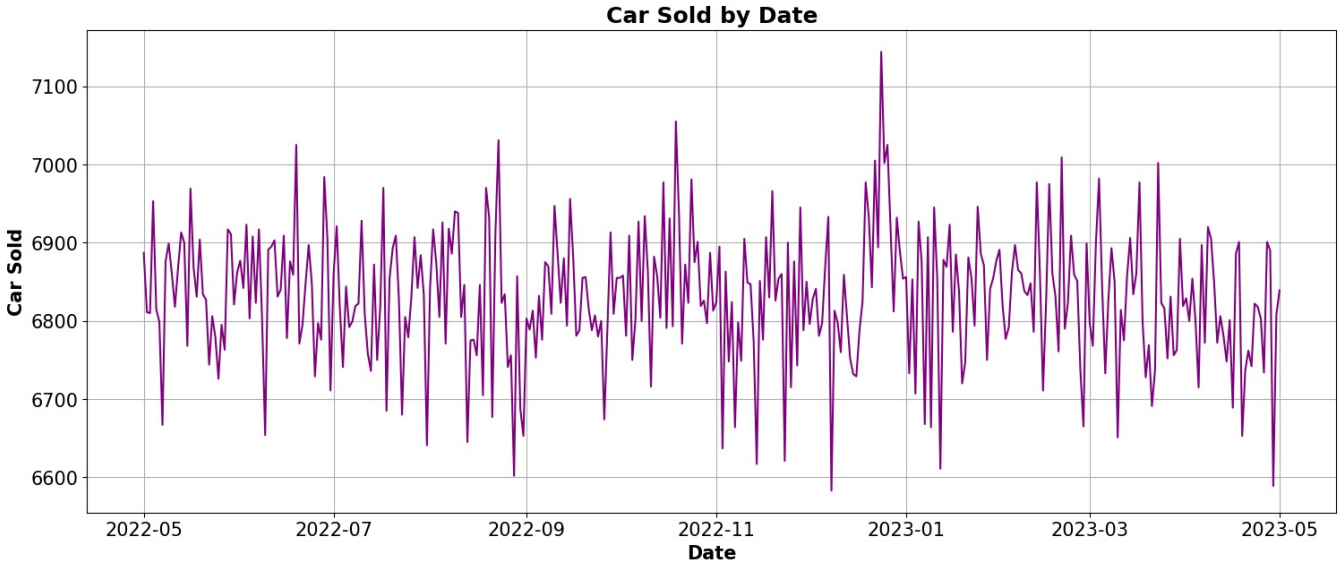
	Date	Car Sold
0	2022-05-01	6887
1	2022-05-02	6811
2	2022-05-03	6810
3	2022-05-04	6953
4	2022-05-05	6815
5	2022-05-06	6799
6	2022-05-07	6667
7	2022-05-08	6876
8	2022-05-09	6899
9	2022-05-10	6861

In [251…
```
Sales_by_Date["Car Sold"].sum()
```

Out[251]: 2500000

In [252…
```
plt.figure(figsize =(18,7))
sns.lineplot(x = "Date", y = "Car Sold", data = Sales_by_Date, color ="purple")
plt.title("Car Sold by Date", weight = "bold", size = 18)
```

```
plt.xticks(size =15)
plt.yticks(size =15)
plt.xlabel("Date", weight = "bold", size =15)
plt.ylabel("Car Sold", weight = "bold", size =15)
plt.grid()
plt.show()
```



In [253…
```
df['Month'] = df['Date'].dt.month
```

In [254…
```
df
```

Out[254]:

	Date	Salesperson	Customer Name	Car Make	Car Model	Car Year	Sale Price	Commission Rate	Commission Earned	Sale Year	Month
0	2022-08-01	Monica Moore MD	Mary Butler	Nissan	Altima	2018	15983	0.070495	1126.73	2022	8
1	2023-03-15	Roberto Rose	Richard Pierce	Nissan	F-150	2016	38474	0.134439	5172.40	2023	3
2	2023-04-29	Ashley Ramos	Sandra Moore	Ford	Civic	2016	33340	0.114536	3818.63	2023	4
3	2022-09-04	Patrick Harris	Johnny Scott	Ford	Altima	2013	41937	0.092191	3866.20	2022	9
4	2022-06-16	Eric Lopez	Vanessa Jones	Honda	Silverado	2022	20256	0.113490	2298.85	2022	6
...	...	...	...	...	...	...	...	...	...	...	...
2499995	2022-05-26	Isabella Moore	Shirley Lee	Chevrolet	Silverado	2021	49823	0.062977	3137.70	2022	5
2499996	2022-10-03	Kimberly Snow	Tara Rodgers	Ford	F-150	2022	18803	0.068339	1284.97	2022	10
2499997	2022-06-07	Jessica Young	Jennifer Moore	Chevrolet	Civic	2010	30863	0.088915	2744.19	2022	6
2499998	2023-02-15	Donald Barber	Ashley Diaz	Honda	Silverado	2014	26125	0.088260	2305.80	2023	2
2499999	2023-03-24	Kayla Fowler	Nathan Thompson	Honda	Civic	2010	20762	0.137105	2846.57	2023	3

2500000 rows × 11 columns

In [255…
```
CSBM = df.groupby("Month")["Month"].agg("count").to_frame()
CSBM.columns = ["Car Sales"]
CSBM = CSBM.reset_index()
CSBM
```

Out[255]:

	Month	Car Sales
0	1	211693
1	2	191524
2	3	211459
3	4	203983
4	5	218840
5	6	205421
6	7	211428
7	8	211446
8	9	204798
9	10	212586
10	11	204302
11	12	212520

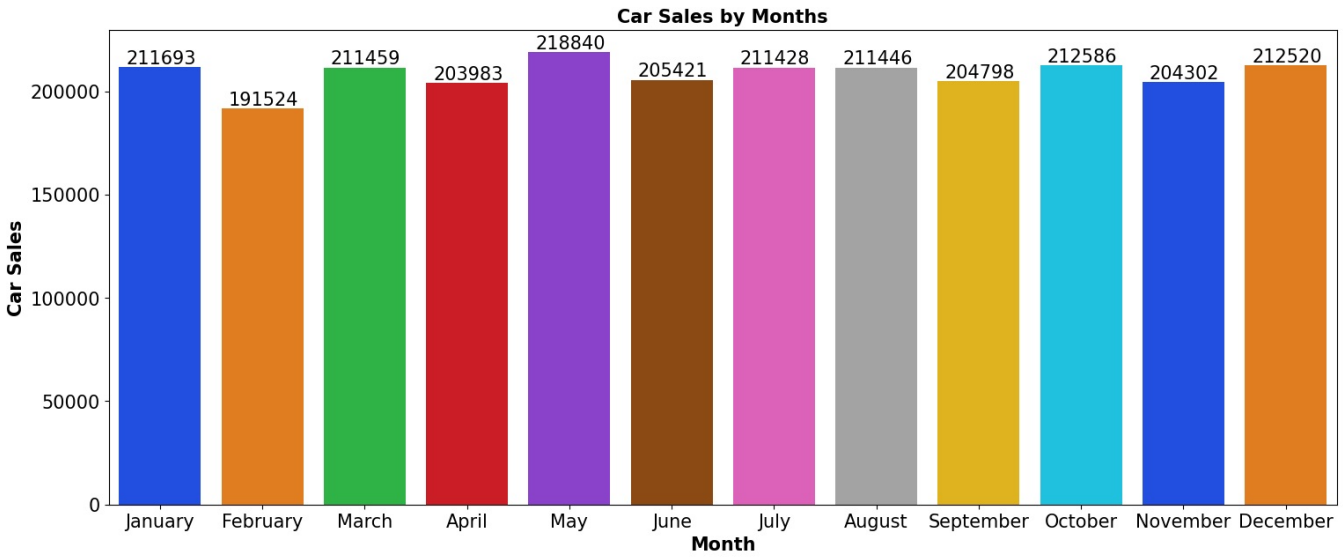
In [256…
```
CSBM['Month'] = CSBM['Month'].apply(lambda x: calendar.month_name[x])
```

In [257…
```
CSBM
```

Out[257]:

	Month	Car Sales
0	January	211693
1	February	191524
2	March	211459
3	April	203983
4	May	218840
5	June	205421
6	July	211428
7	August	211446
8	September	204798
9	October	212586
10	November	204302
11	December	212520

In [258…
```
plt.figure(figsize =(18, 7))
graph =sns.barplot(x ="Month", y ="Car Sales", data = CSBM, palette ="bright")
for p in graph.patches:
        graph.annotate('{:.0f}'.format(p.get_height()),
                       (p.get_x()+0.41, p.get_height()),
                        ha='center', va='bottom',color= 'black', size = 15)
plt.title("Car Sales by Months", weight ="bold", size =15)
plt.xticks(size =15)
plt.yticks(size =15)
plt.xlabel("Month", weight ="bold", size =15)
plt.ylabel("Car Sales", weight ="bold", size =15)
plt.show()
```



In [259…
```
date = df[["Date"]]
date.head(1)
```

Out[259]:

	Date
0	2022-08-01

In [260…
```
DT =date.groupby("Date")["Date"].agg("count").to_frame()
DT.columns =["Total Sales"]
DT = DT.reset_index()
DT
```

Out[260]:

	Date	Total Sales
0	2022-05-01	6887
1	2022-05-02	6811
2	2022-05-03	6810
3	2022-05-04	6953
4	2022-05-05	6815
...	...	...
361	2023-04-27	6901
362	2023-04-28	6890
363	2023-04-29	6589
364	2023-04-30	6808
365	2023-05-01	6839

366 rows × 2 columns

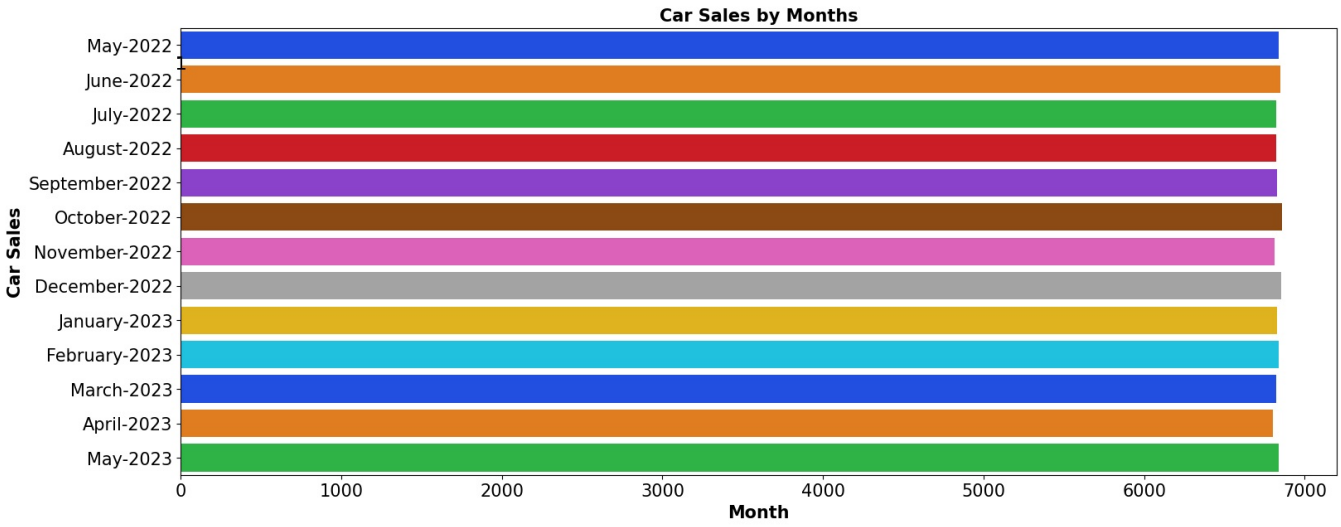
In [261…
```
DT["Month_Year"] = DT["Date"].dt.strftime("%B-%Y")
DT
```

Out[261]:

	Date	Total Sales	Month_Year
0	2022-05-01	6887	May-2022
1	2022-05-02	6811	May-2022
2	2022-05-03	6810	May-2022
3	2022-05-04	6953	May-2022
4	2022-05-05	6815	May-2022
...	...	...	...
361	2023-04-27	6901	April-2023
362	2023-04-28	6890	April-2023
363	2023-04-29	6589	April-2023
364	2023-04-30	6808	April-2023
365	2023-05-01	6839	May-2023

366 rows × 3 columns

In [265…
```
plt.figure(figsize =(18, 7))
graph =sns.barplot(y ="Month_Year", x ="Total Sales", data = DT, palette ="bright", errorbar =None)
for p in graph.patches:
        graph.annotate('{:.0f}'.format(p.get_height()),
                       (p.get_x()+0.41, p.get_height()),
                        ha='center', va='bottom',color= 'black', size = 15)
plt.title("Car Sales by Months", weight ="bold", size =15)
plt.xticks(size =15)
plt.yticks(size =15)
plt.xlabel("Month", weight ="bold", size =15)
plt.ylabel("Car Sales", weight ="bold", size =15)
plt.show()
```



In [269…
```
del(df["Month"])
```

In [270…
```
df
```

Out[270]:

	Date	Salesperson	Customer Name	Car Make	Car Model	Car Year	Sale Price	Commission Rate	Commission Earned	Sale Year
0	2022-08-01	Monica Moore MD	Mary Butler	Nissan	Altima	2018	15983	0.070495	1126.73	2022
1	2023-03-15	Roberto Rose	Richard Pierce	Nissan	F-150	2016	38474	0.134439	5172.40	2023
2	2023-04-29	Ashley Ramos	Sandra Moore	Ford	Civic	2016	33340	0.114536	3818.63	2023
3	2022-09-04	Patrick Harris	Johnny Scott	Ford	Altima	2013	41937	0.092191	3866.20	2022
4	2022-06-16	Eric Lopez	Vanessa Jones	Honda	Silverado	2022	20256	0.113490	2298.85	2022
...	...	...	...	...	...	...	...	...	...	...
2499995	2022-05-26	Isabella Moore	Shirley Lee	Chevrolet	Silverado	2021	49823	0.062977	3137.70	2022
2499996	2022-10-03	Kimberly Snow	Tara Rodgers	Ford	F-150	2022	18803	0.068339	1284.97	2022
2499997	2022-06-07	Jessica Young	Jennifer Moore	Chevrolet	Civic	2010	30863	0.088915	2744.19	2022
2499998	2023-02-15	Donald Barber	Ashley Diaz	Honda	Silverado	2014	26125	0.088260	2305.80	2023
2499999	2023-03-24	Kayla Fowler	Nathan Thompson	Honda	Civic	2010	20762	0.137105	2846.57	2023

2500000 rows × 10 columns

In [274…
```
df["Month_year"] =df["Date"].dt.strftime("%B-%Y")
df
```

Out[274]:

	Date	Salesperson	Customer Name	Car Make	Car Model	Car Year	Sale Price	Commission Rate	Commission Earned	Sale Year	Month_year
0	2022-08-01	Monica Moore MD	Mary Butler	Nissan	Altima	2018	15983	0.070495	1126.73	2022	August-2022
1	2023-03-15	Roberto Rose	Richard Pierce	Nissan	F-150	2016	38474	0.134439	5172.40	2023	March-2023
2	2023-04-29	Ashley Ramos	Sandra Moore	Ford	Civic	2016	33340	0.114536	3818.63	2023	April-2023
3	2022-09-04	Patrick Harris	Johnny Scott	Ford	Altima	2013	41937	0.092191	3866.20	2022	September-2022
4	2022-06-16	Eric Lopez	Vanessa Jones	Honda	Silverado	2022	20256	0.113490	2298.85	2022	June-2022
...	...	...	...	...	...	...	...	...	...	...	...
2499995	2022-05-26	Isabella Moore	Shirley Lee	Chevrolet	Silverado	2021	49823	0.062977	3137.70	2022	May-2022
2499996	2022-10-03	Kimberly Snow	Tara Rodgers	Ford	F-150	2022	18803	0.068339	1284.97	2022	October-2022
2499997	2022-06-07	Jessica Young	Jennifer Moore	Chevrolet	Civic	2010	30863	0.088915	2744.19	2022	June-2022
2499998	2023-02-15	Donald Barber	Ashley Diaz	Honda	Silverado	2014	26125	0.088260	2305.80	2023	February-2023
2499999	2023-03-24	Kayla Fowler	Nathan Thompson	Honda	Civic	2010	20762	0.137105	2846.57	2023	March-2023

2500000 rows × 11 columns

In [281…

```
CEMS = df[(df["Salesperson"] =="Michael Smith") & (df["Car Make"] =="Nissan") & (df["Month_year"] =="December-2
CEMS
```

Out[281]:

	Date	Salesperson	Customer Name	Car Make	Car Model	Car Year	Sale Price	Commission Rate	Commission Earned	Sale Year	Month_year
204674	2022-12-15	Michael Smith	Carolyn Russell	Nissan	Corolla	2014	30256	0.122109	3694.52	2022	December-2022
205491	2022-12-24	Michael Smith	Tina Rogers	Nissan	Altima	2012	25760	0.053259	1371.96	2022	December-2022
292009	2022-12-03	Michael Smith	Manuel Williams	Nissan	Civic	2015	13427	0.063041	846.45	2022	December-2022
398112	2022-12-30	Michael Smith	Joshua Clark	Nissan	F-150	2020	10261	0.127373	1306.97	2022	December-2022
517166	2022-12-17	Michael Smith	Kellie Williams	Nissan	F-150	2022	48221	0.141085	6803.28	2022	December-2022
740258	2022-12-23	Michael Smith	Courtney Jones	Nissan	F-150	2012	29463	0.104427	3076.73	2022	December-2022
863920	2022-12-12	Michael Smith	Jasmine Taylor	Nissan	Corolla	2021	10485	0.071749	752.29	2022	December-2022
986503	2022-12-15	Michael Smith	Christina Patel	Nissan	Altima	2021	26360	0.070925	1869.58	2022	December-2022
1028644	2022-12-14	Michael Smith	Traci Robinson	Nissan	Silverado	2012	20895	0.128342	2681.71	2022	December-2022
1058318	2022-12-03	Michael Smith	Alison Anderson	Nissan	Silverado	2022	40817	0.092968	3794.66	2022	December-2022
1107097	2022-12-18	Michael Smith	Shawn Morrison	Nissan	Altima	2018	16779	0.129140	2166.85	2022	December-2022
1119870	2022-12-27	Michael Smith	Stacy Lee	Nissan	Altima	2012	10608	0.129892	1377.90	2022	December-2022
1566975	2022-12-17	Michael Smith	Maria Carter	Nissan	F-150	2013	33218	0.079050	2625.87	2022	December-2022
1767250	2022-12-20	Michael Smith	Mary Silva	Nissan	Silverado	2013	12853	0.082192	1056.41	2022	December-2022
1807044	2022-12-14	Michael Smith	Michelle Wilson	Nissan	Corolla	2011	30832	0.121499	3746.05	2022	December-2022
1972272	2022-12-15	Michael Smith	Thomas Ross	Nissan	Corolla	2020	11569	0.069995	809.77	2022	December-2022
2439535	2022-12-22	Michael Smith	Derrick Jenkins	Nissan	Altima	2011	48348	0.119158	5761.04	2022	December-2022

In [286…

```
Commission_Earned = CEMS["Commission Earned"].agg("sum")
print("Total Commission Earned By Micheal Smith in December 2022 = $" +str(format(Commission_Earned, ",")))
```

```
Total Commission Earned By Micheal Smith in December 2022 = $43,742.04
```