

# Projects Report

Made By Arnav Kumar Shukla  
(<https://github.com/ARNAVKS>)

# About the Projects

In this 1 month Data Science internship from Unified Mentor. I have done 2 projects in it, i have done the coding part and now i am making reports of both the project of the resut and listing some of my finding and analysis from both of the projects.

## **Analysis of Amazon Sales Data**

**01.**

Analyze the sales trend of the given Amazon sales data from 2010 to 2017. Do monthly, yearly and month-yearly analysis and comes up with your finding from it.

## **Data Visualization of Bird Strikes**

**02.**

Analyze the Bird Strike Data from 2000 to 2011. After Analyzing find and solve the given case studies.

# **Analysis of Amazon Sales Data**

# Quantitative Information got from the Dataset

- Total Revenue generated from 2010 to 2017 : 137348768.31
- Total Profit generated from 2010 to 2017 : 44168198.40
- Total Cost from 2010 to 2017 : 93180569.91
- Average Unit Cost : 191.05
- Average Unit Price : 276.76
- Average Revenue generated per order : 1373487.68
- Average Profit got per order : 441681.98
- Average Cost per order : 931805.70
- Total Units Sold from 2010 to 2017 : 512871
- Average Units sold per order : 5128.71

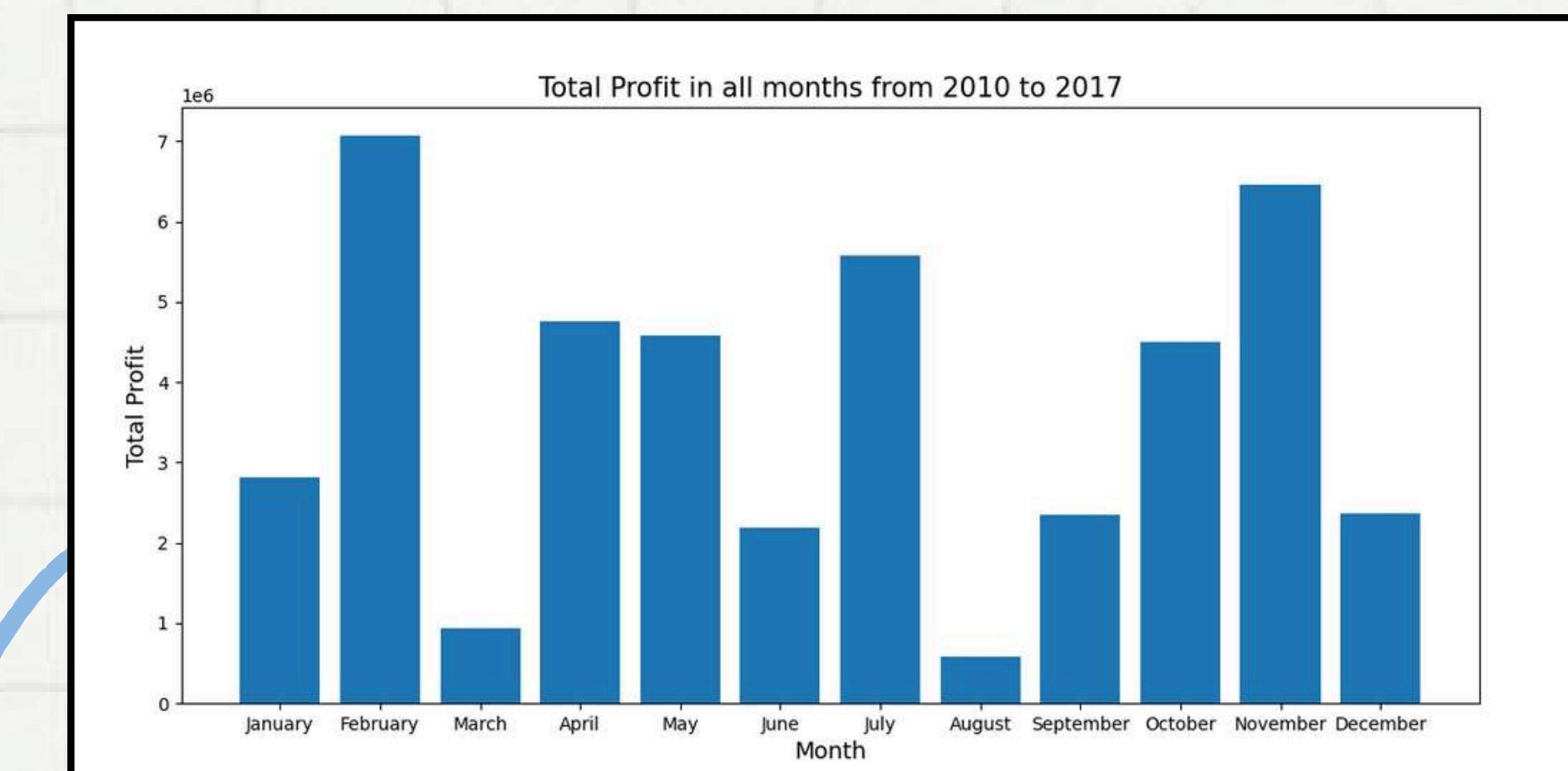
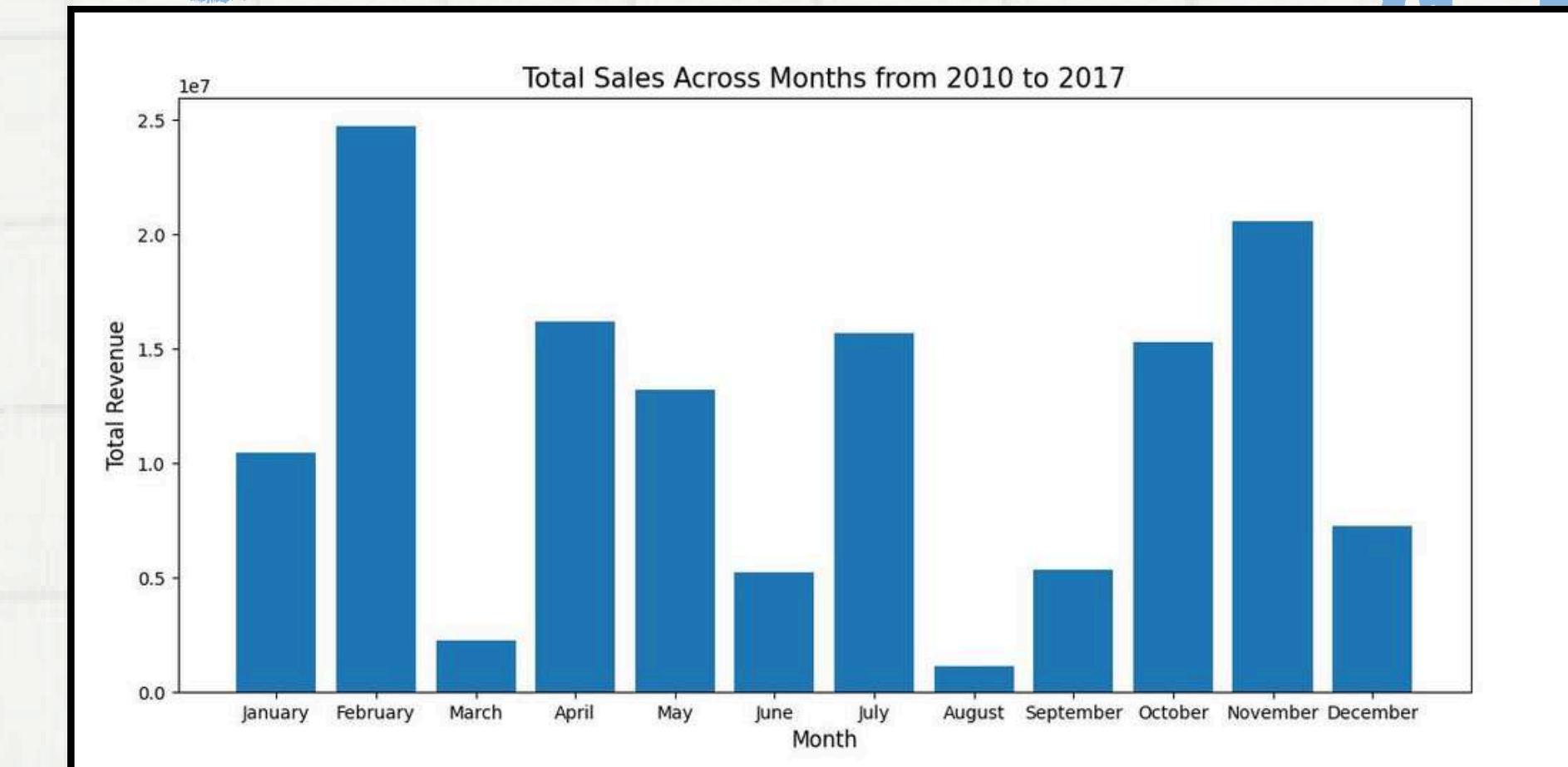
# Monthly Sales

This is the monthly sales bar graph from 2010 to 2017. We got a result that we have more sales in February, April, May, July, October and November and the least sales in March and August and others have normal sales. Sales can be constant in every month but there are some increment and decrement in the total sales.

The Reason is :

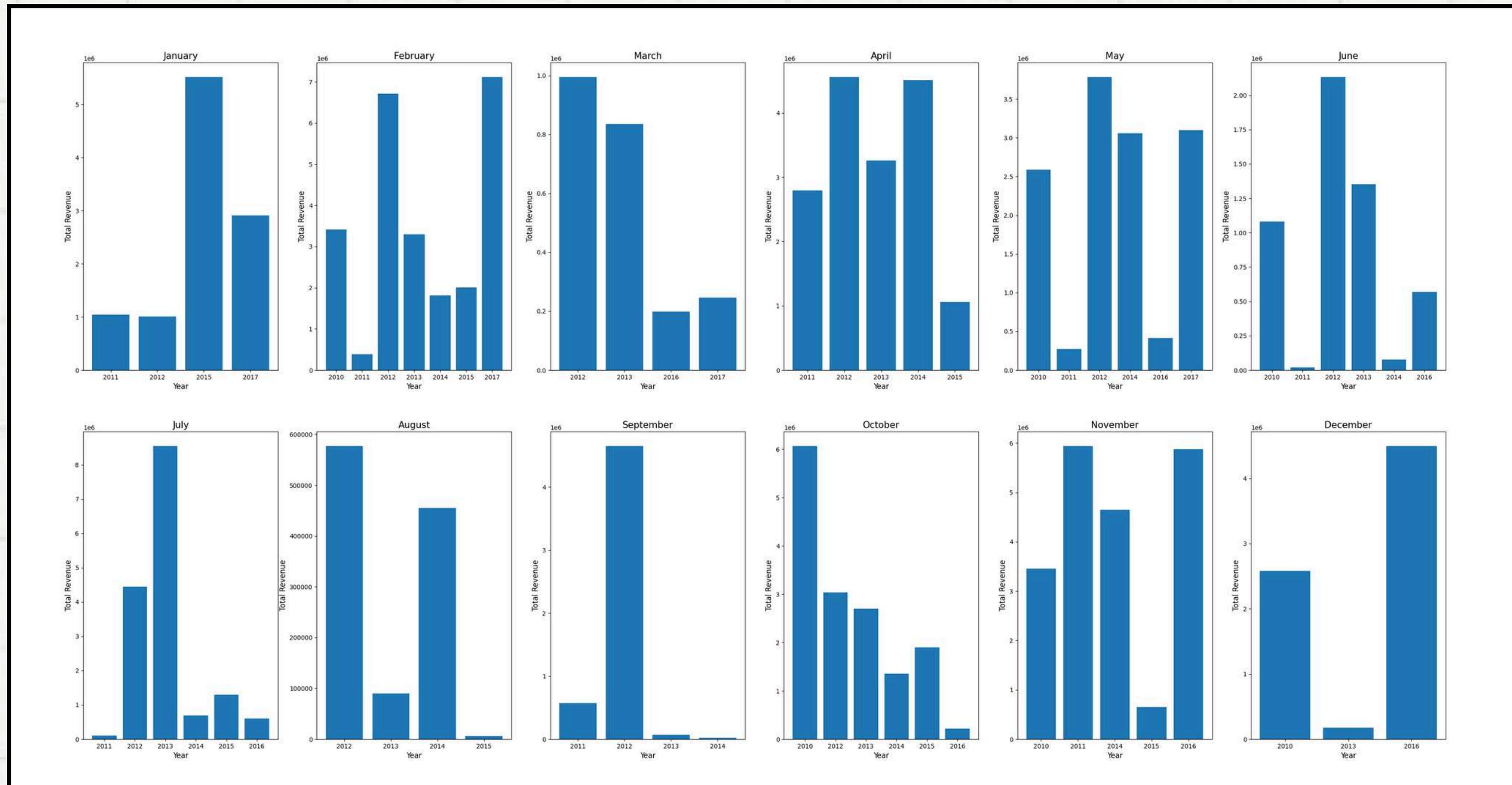
- For High Sales
  1. **February** – It can happen due to Valentine's Days people gives gift to their loved ones, others reason probably winter sales.
  2. **April** – It can happen due to Spring Season.
  3. **May** – It can happen because this a graduation month in most of the country, many people graduates many people congratulates them with gifts.
  4. **July** – It can happen due to Amazon Prime Day.
  5. **October** – It can happen due to Halloween.
  6. **November** – It can happen due to Pre-Holiday Shopping and also high sales due to Black Friday.
- For Low Sales
  1. **March** – It is generally a very working month.
  2. **August** – it is due to end of Summer Holiday

You can also see another graph of the Total Profit, it looked approximately similar to the Total Sales. This shows a positive indication for the finance, it means there not much loss has been taken throughout all years.



# Graph of All Months over 2010 to 2017

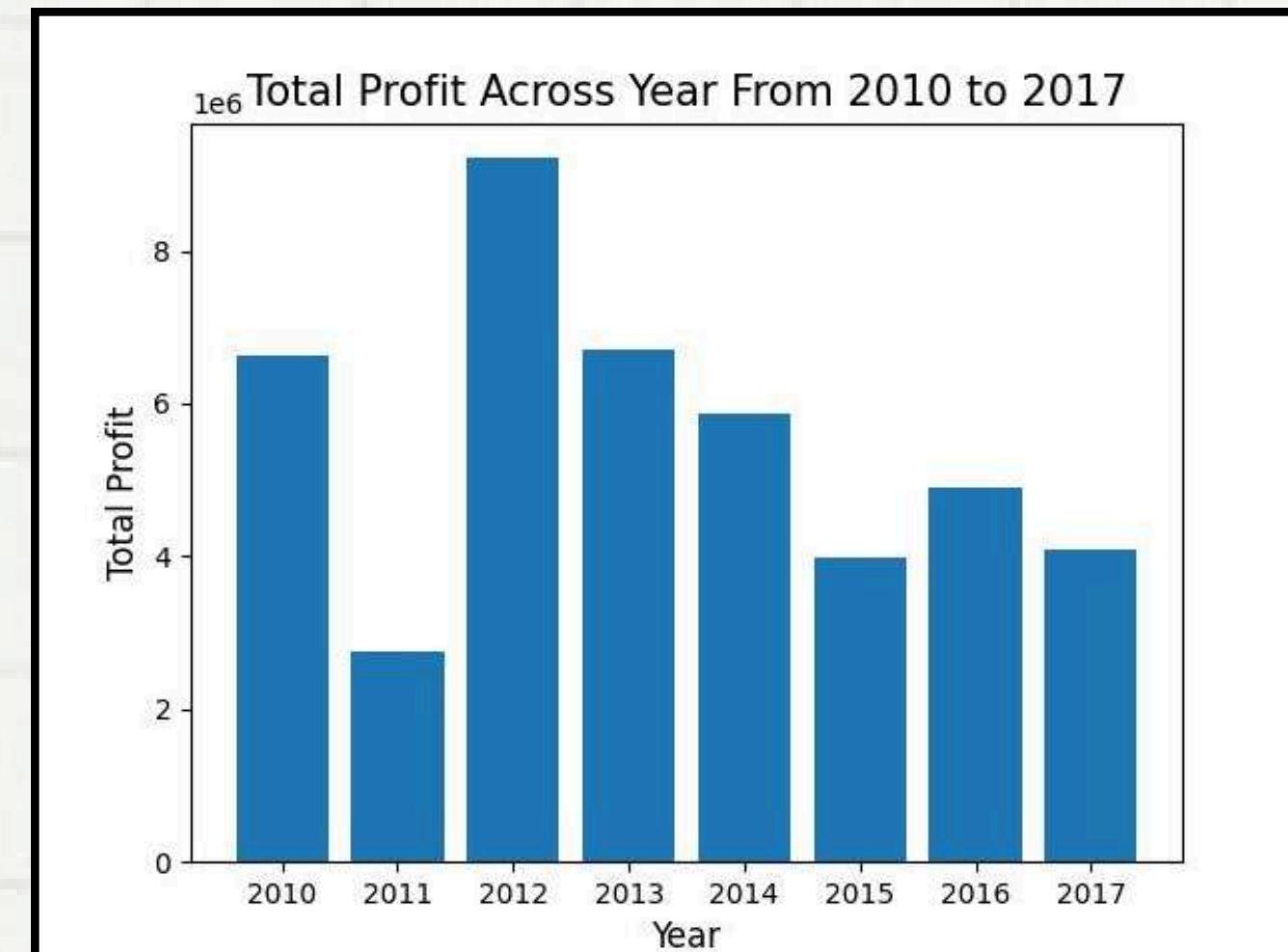
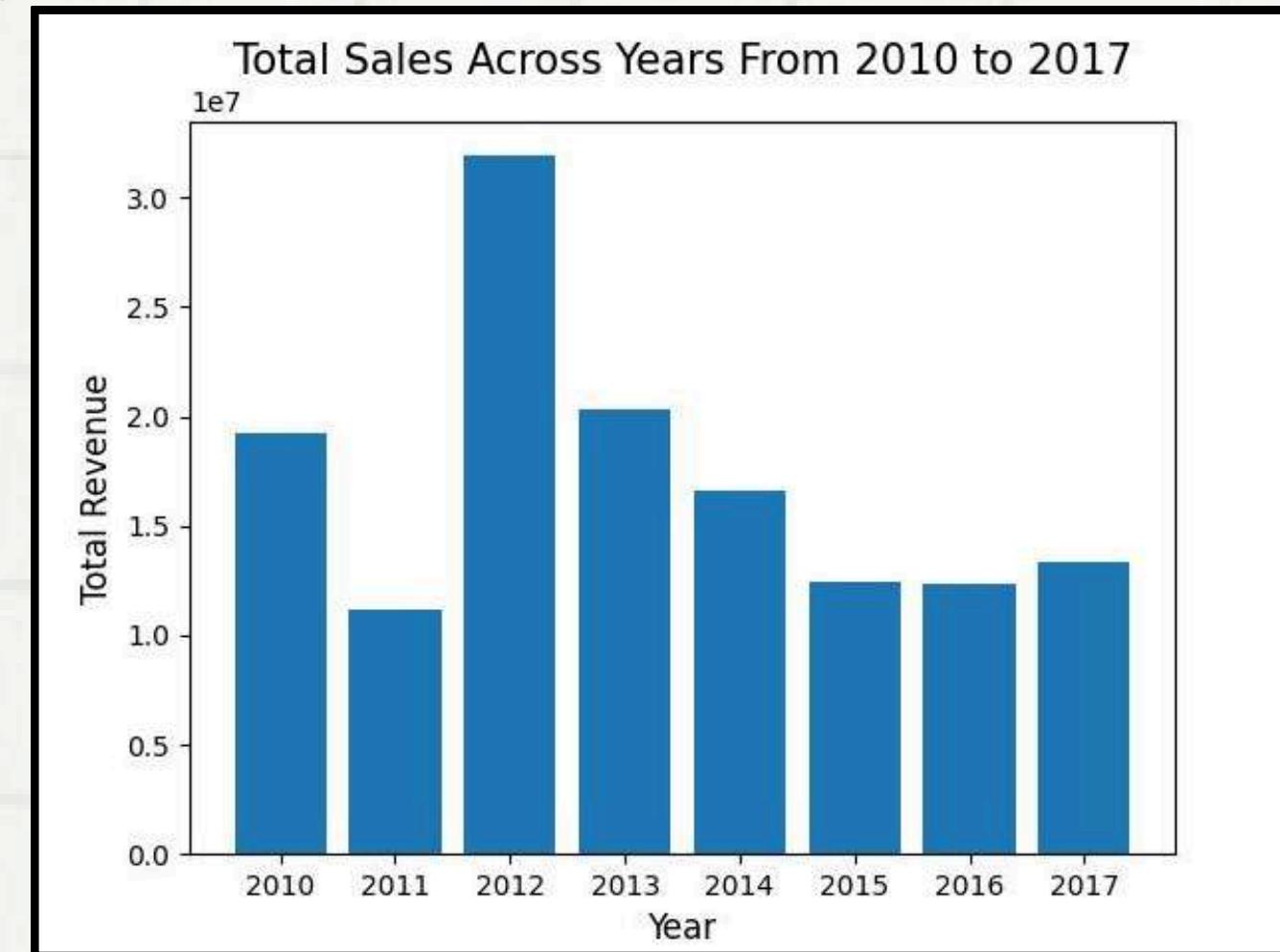
(Note-If any year was not mentioned in graph of a particular month then there were no sales in the month of that year)



# Yearly Sales

This is the graph of the Total Yearly Sales from 2010 to 2017. This graph show that 2012 was the year that has the highest sales and 2011 was the year that has the lowest sales. For all year we can say that first in 2010 there were 19186k revenue sales, then in 2011 there was a decrement and have 11129k revenue sales, then in 2012 there was a massive sales increment of almost 31898k revenue sales, in 2013 there was a decrement and got a sales of 20330k revenue, in 2014 there was a slight decrement as compared to 2013 and got sales of 16630k revenue and then in all month after 2014 (i.e. 2015, 2016 and 2017) have almost similiar sales, approximately of 12500k revenue in 2015, 2016 and 2017.

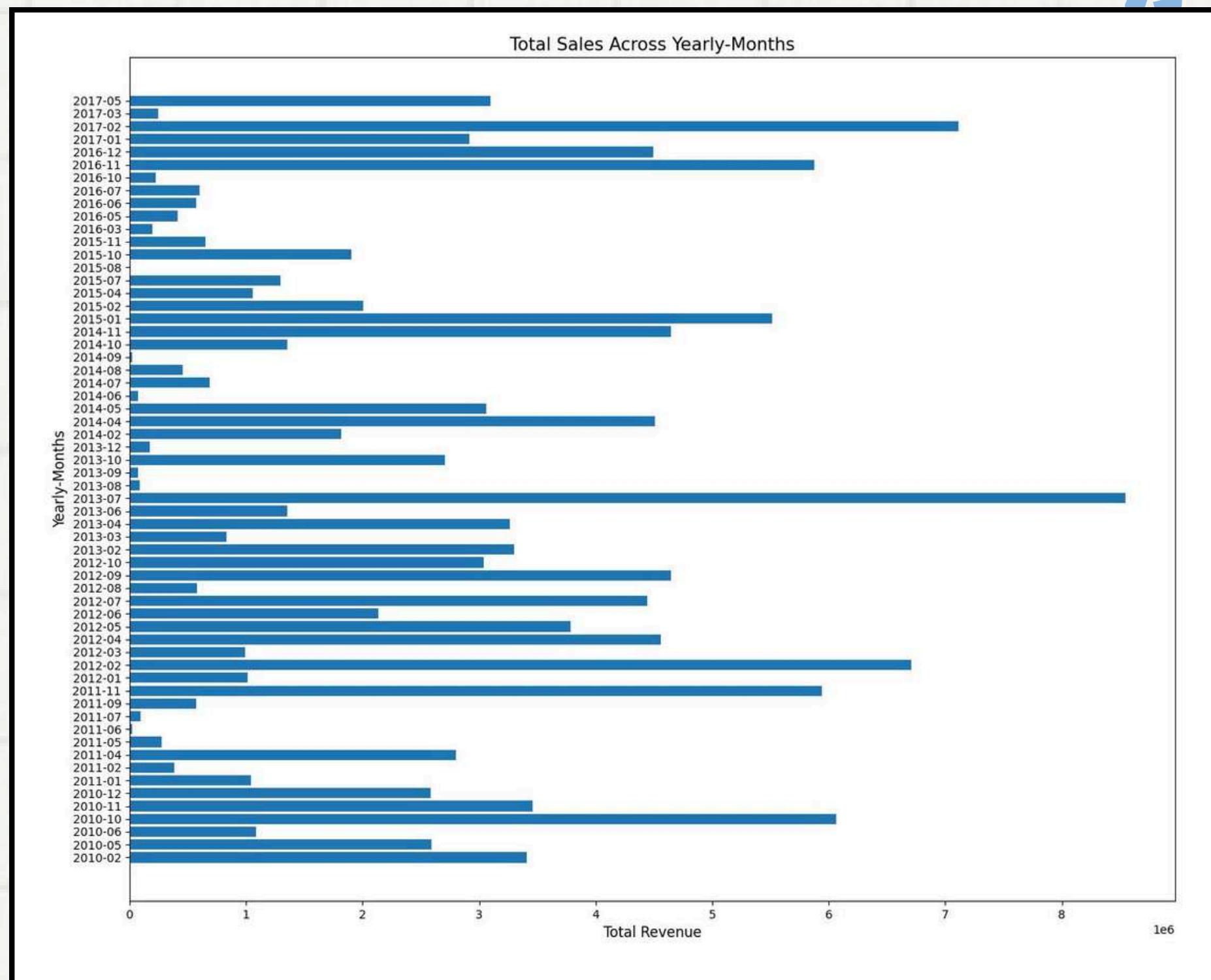
You can also see another graph of the Total Profit, it looked approximately similiar to the Total Sales. This shows a positive indication for the finance, it mean there not any significant loss is taken by amazon throughout the year 2010 to 2017.



# Yearly-Month Sales

This is the yearly-month sales graph. This graph shows that there were highest sales in July month of the year 2013 and there were lowest sales in August month of the year of 2015.

- **Average Yearly-Month Sales** – 2289146.1385
- **Maximum sales were in July 2013 : 8545511.2**
- **Minimum sales were in August 2015 : 6279.09**

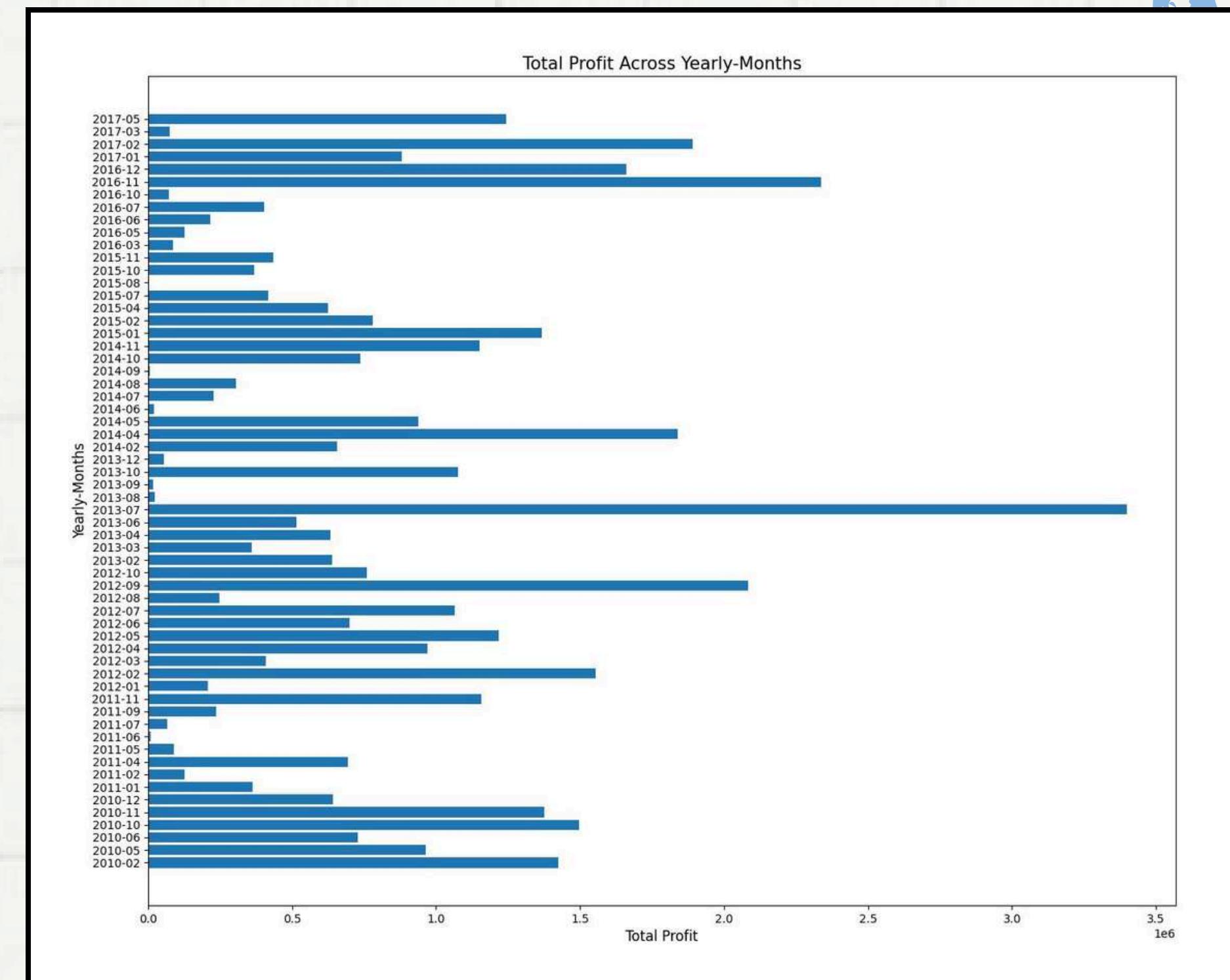


# Yearly-Month Profit

This is the yearly-month profit graph. This graph shows that there were highest profit in July month of the year 2013 and there were lowest profit in August month of the year of 2015.

- **Average Yearly-Month Sales – 736136.64**
- **Maximum sales were in July 2013 : 3398463.02**
- **Minimum sales were in August 2015 : 1621.93**

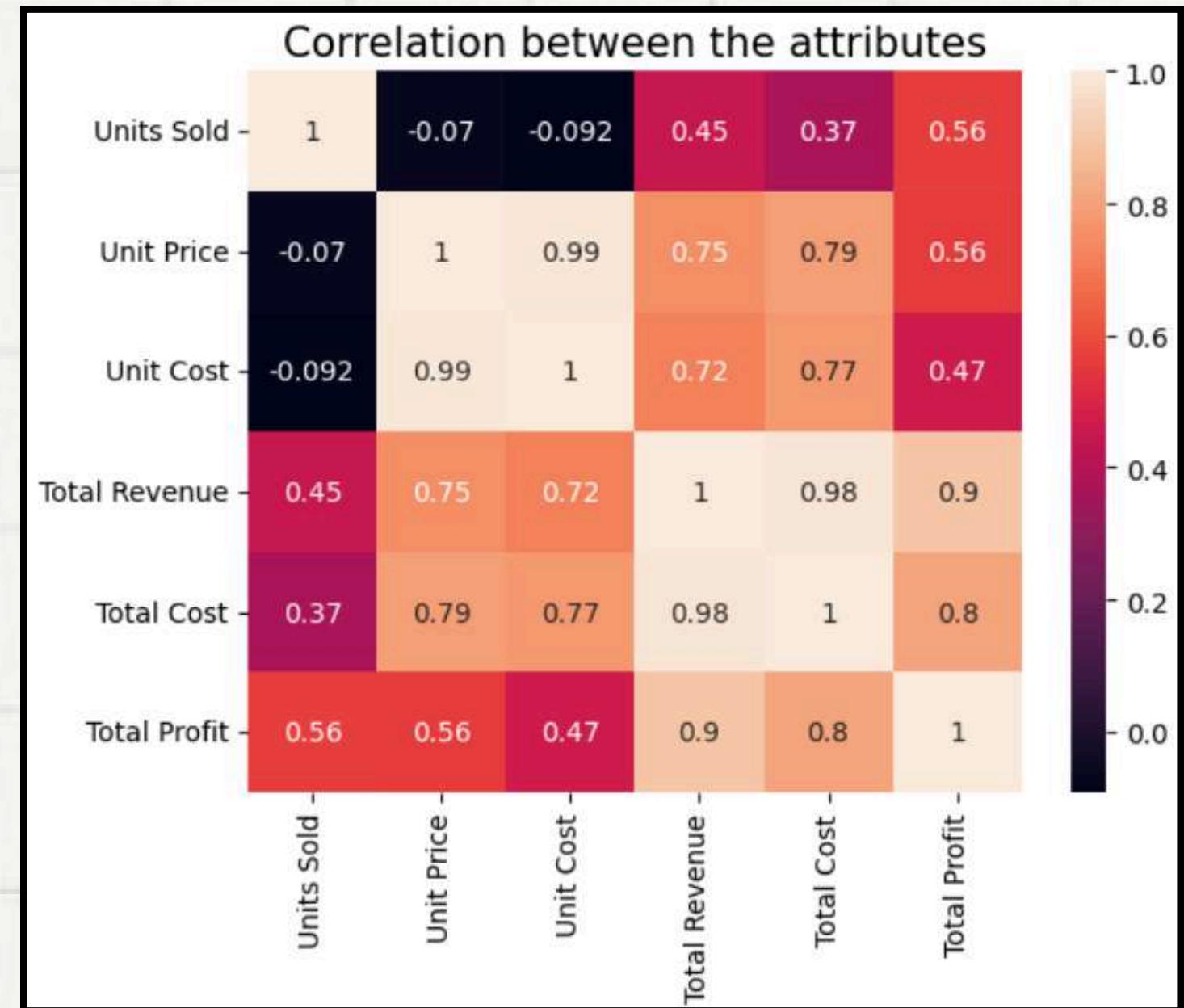
You can see that from the previous yearly-month sales graph we can say that both maximum sales and profit goes in same month and same year, same goes for minimum sales and profit. After comparing both graph you can see some disimilarities between sales and profit graph which shows that there were some significant losses were taken in some cases that later fades away in yearly and monthly data.



# Relationship between the Attributes

This Correlation Matrix Shows the Relationship between all Attributes and its sums up as :

- **Unit Sold** - If Unit Sold value increases then Unit Price and Unit Cost slightly decreases and for Total Revenue, Total Cost and Total Profit it increases when Unit Sold value increases.
- **Unit Price** - If Unit Price value increases then Unit Sold slightly decreases, for Unit Cost, Total Revenue, Total Cost and Total Profit it will increase when Unit Price value increases.
- **Unit Cost** - If Unit Cost value increases the unit sold value slightly decreases, for Unit Price, Total Revenue, Total Cost and Total Profit it will increase when Unit Cost value increases.
- **Total Revenue** - If Total Revenue value increases then all attributes Units Sold, Unit Price, Unit Cost, Total Cost and Total Profit will also increases.
- **Total Cost** - If Total Cost value increases then all attributes Units Sold, Unit Price, Unit Cost, Total Revenue and Total Profit will also increases.
- **Total Profit** - If Total Profit value increases then all attributes Units Sold, Unit Price, Unit Cost, Total Revenue and Total Cost will also increases.

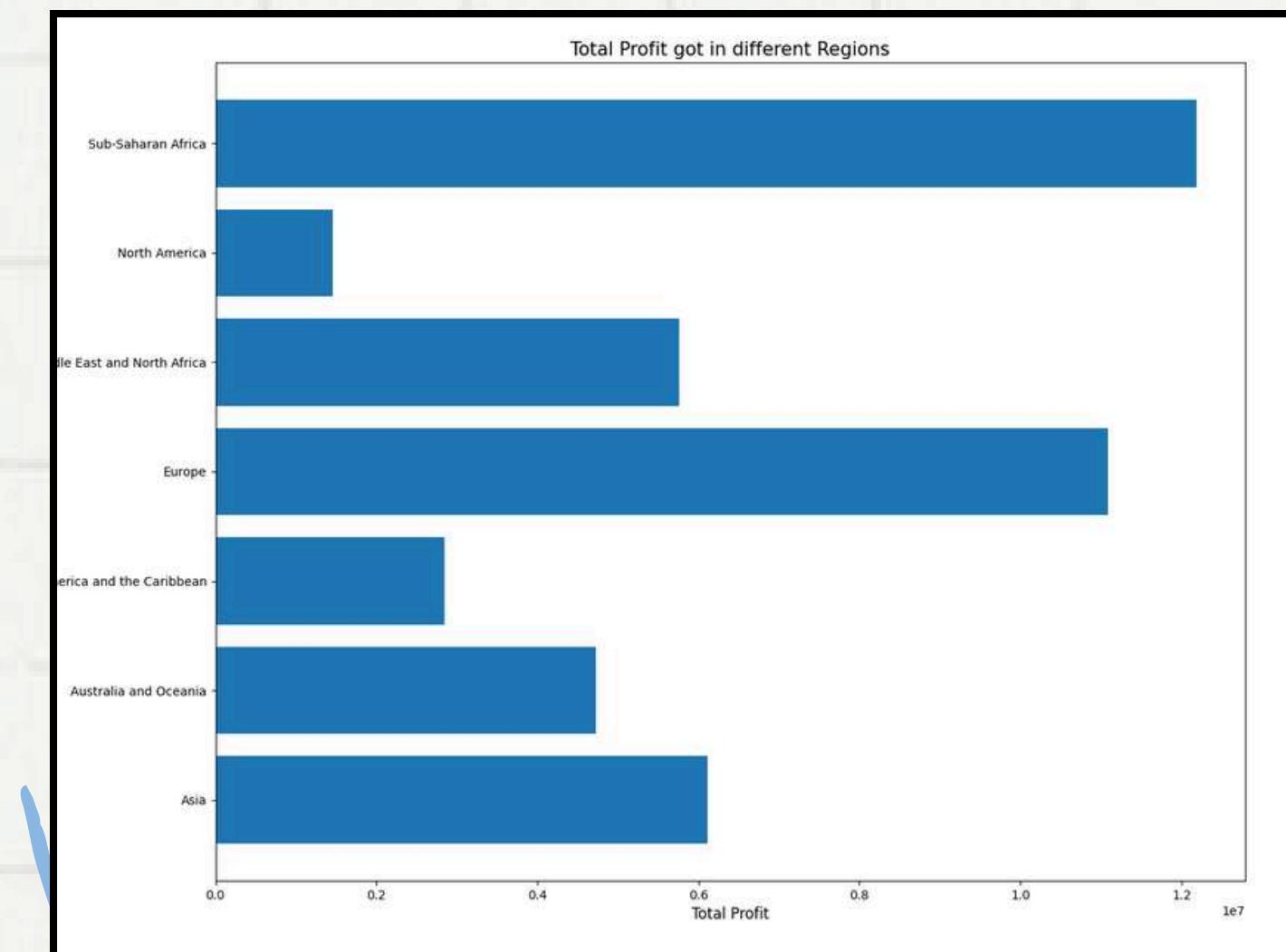
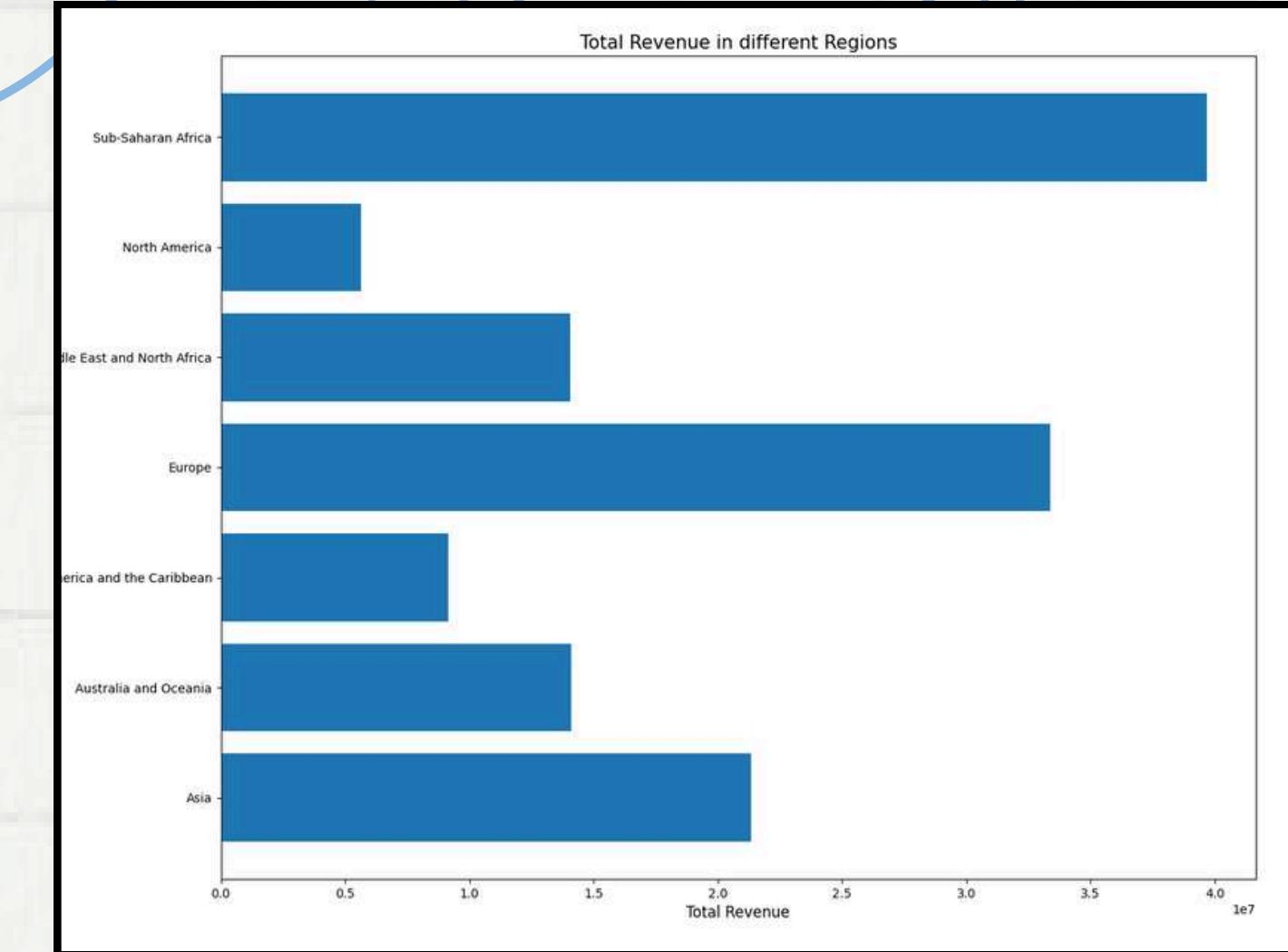


# Region-Wise Analysis

We have different region here, Asia, Australia and Oceania, Central America and the Caribbean, Europe, Middle East and North Africa, North America and Sub-Saharan Africa. We will analyze there sales and profit by plotting their bar graph.

After plotting the graph of the region-wise sales we got as a result that highest sales was from the Sub-Saharan Africa region of 39672031.43 sales and got lowest sales was from the North America of 5643356.55. Europe and Asia also have some decent sales.

Now, as we goes on the profit graph you can see that similarity between Sales graph and Profits graph that concludes that no significant loss has been taken even from the diffrent regions and from the different region they got a sales with great profits..

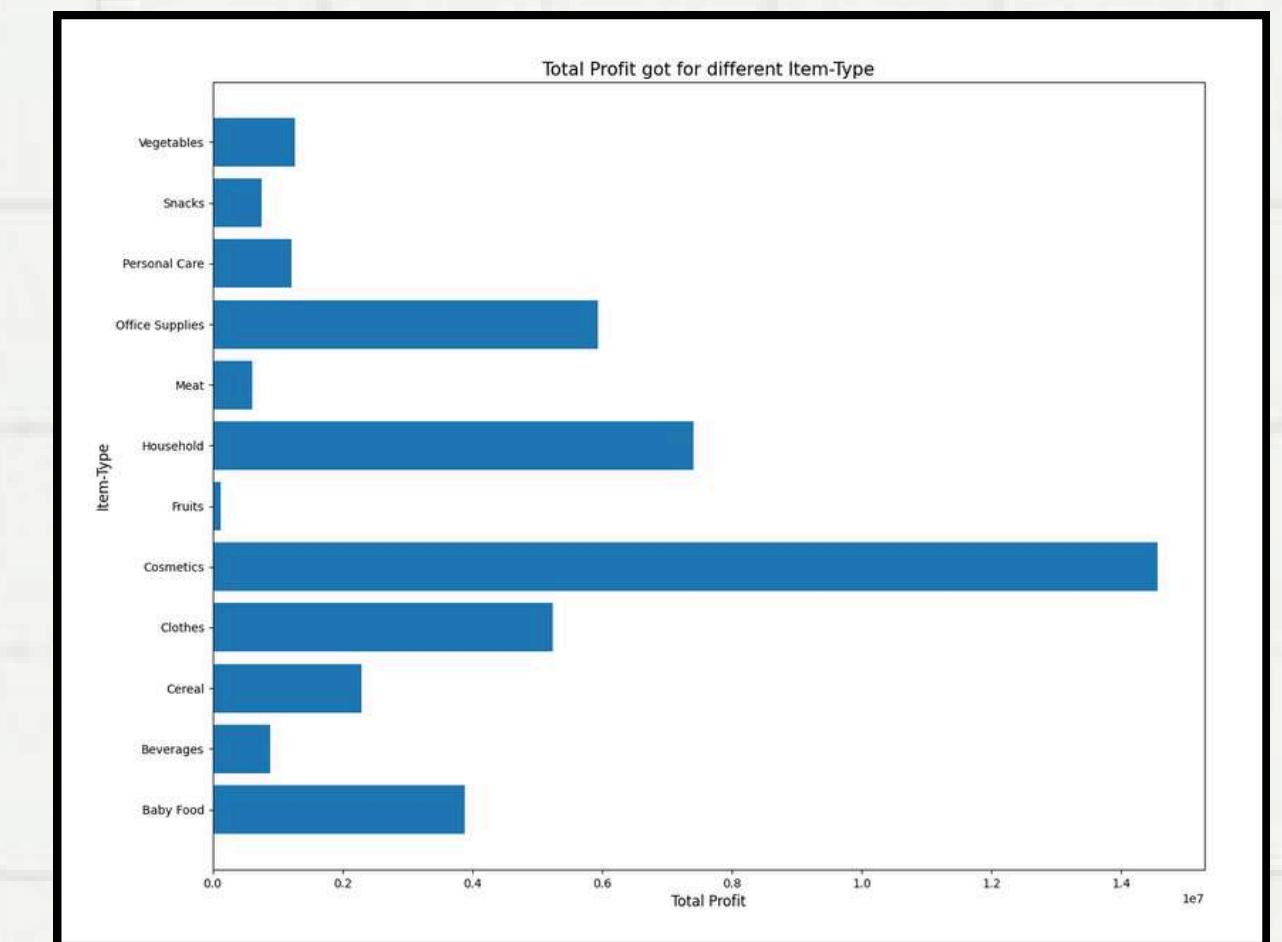
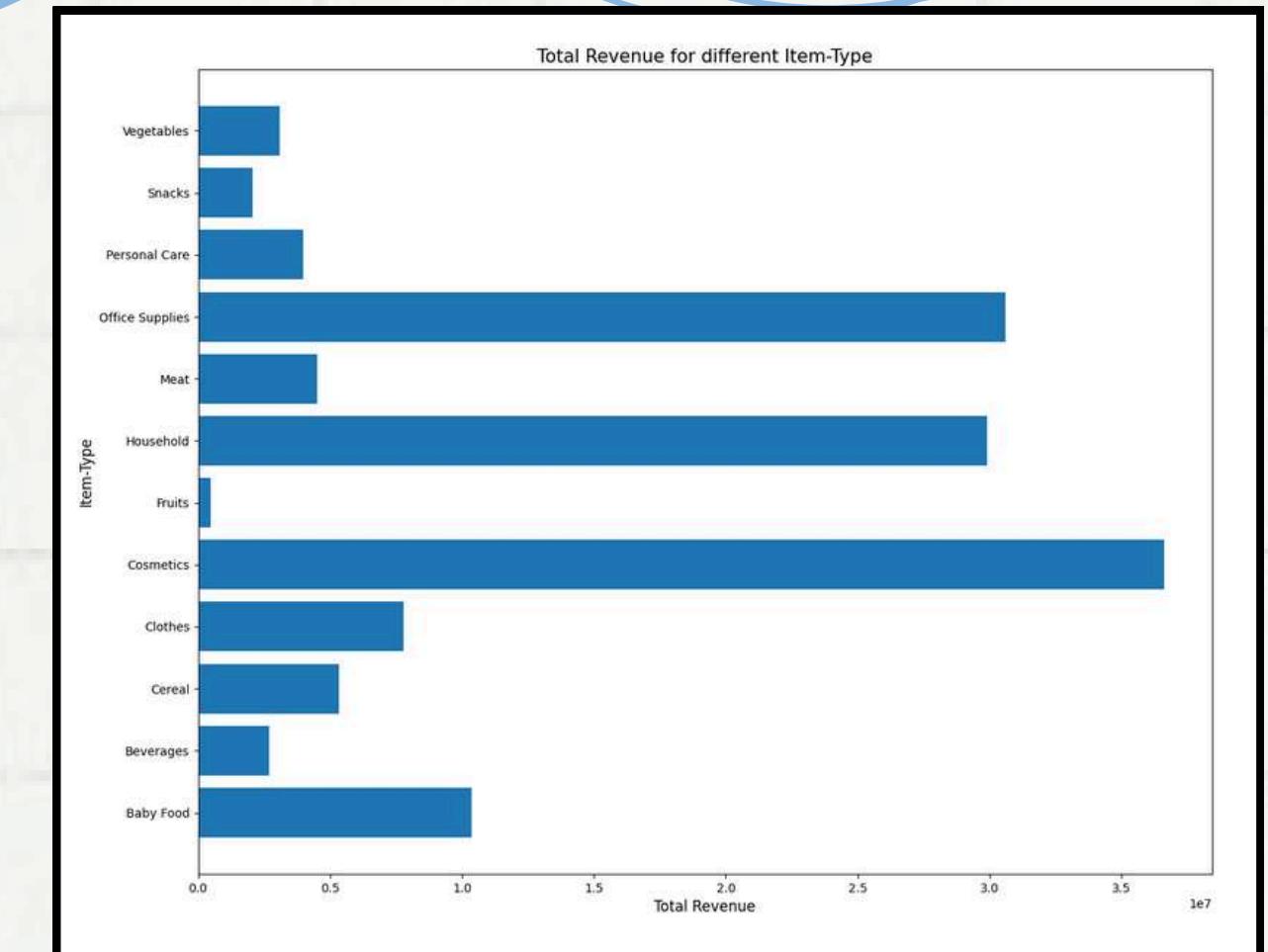


# Item-Type Analysis

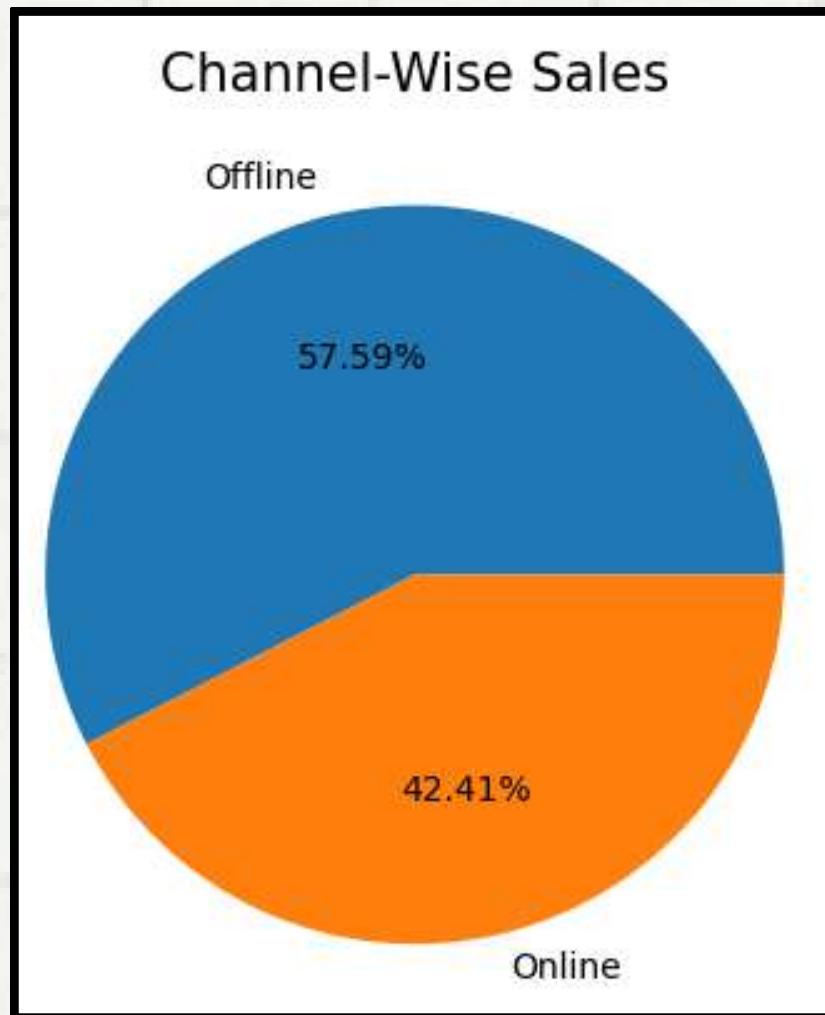
We will analyze the item-type now. There are total 12 item-type :  
Vegetables, Snacks, Personal Care, Office Supplies, Meat, Household,  
Fruits, Cosmetics, Clothes, Cereal, Beverages, Baby Food.

Now take a look at the sales graph which shows that item types such as Office Supplies, Household and Cosmetics got the highest sales and got the lowest sales in Vegetables, Snacks, Fruits and Beverages and then decent sales in all others.

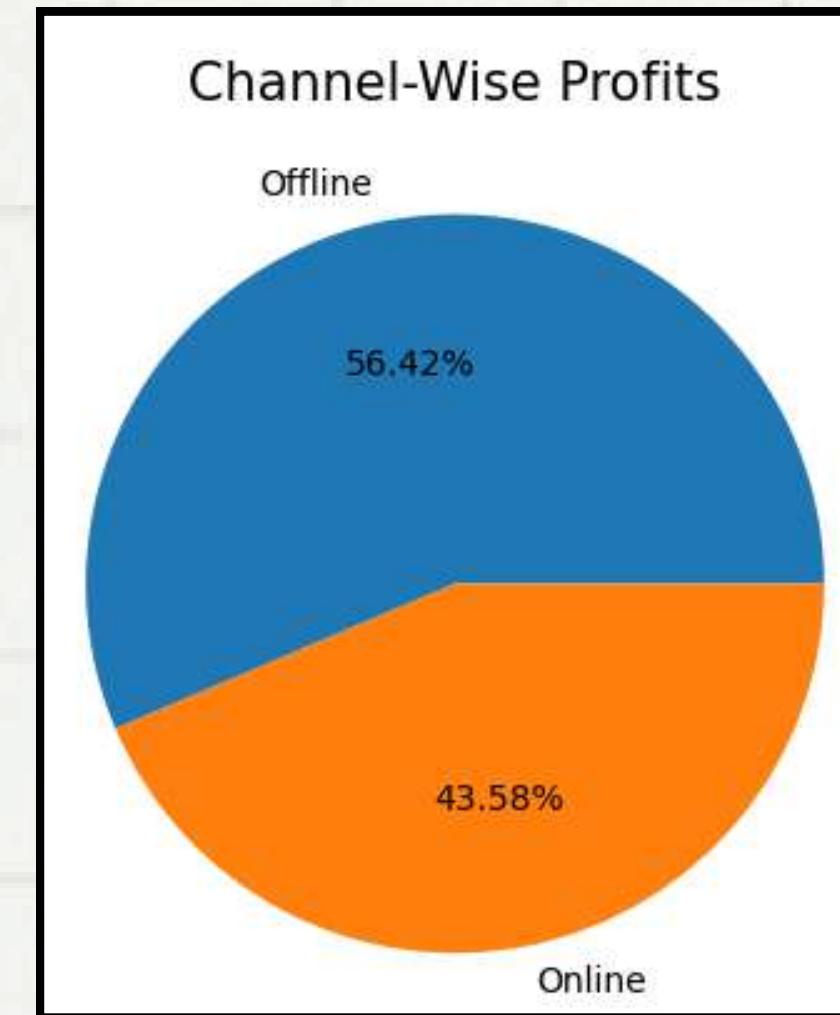
Now take a look at the profits graph which show similar graph for all item-type as compare to sales graph except for the item-type as Office Supplies and Household, this shows that Office Supplies and Household have low profit margin as compare to all others.



# Channel-Wise Analysis



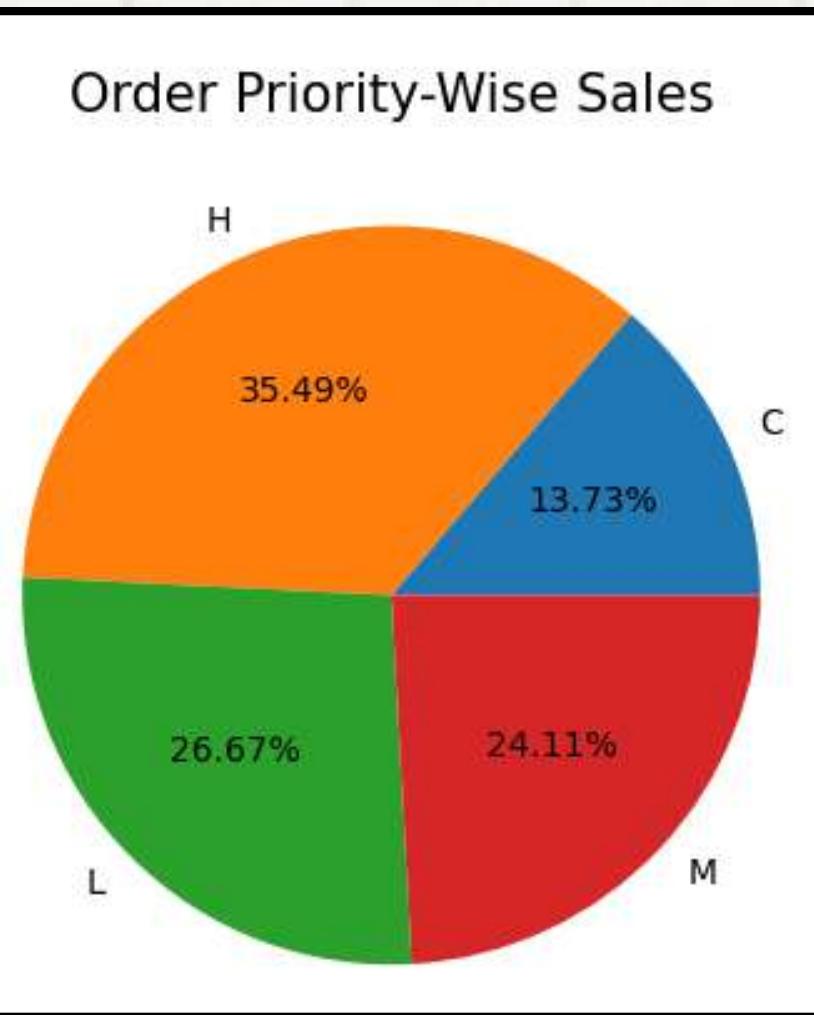
Take a look at the Channel-Wise Sales pie chart which shows that 57.59% Sales were from Offline and 42.41% Sales were from Online.



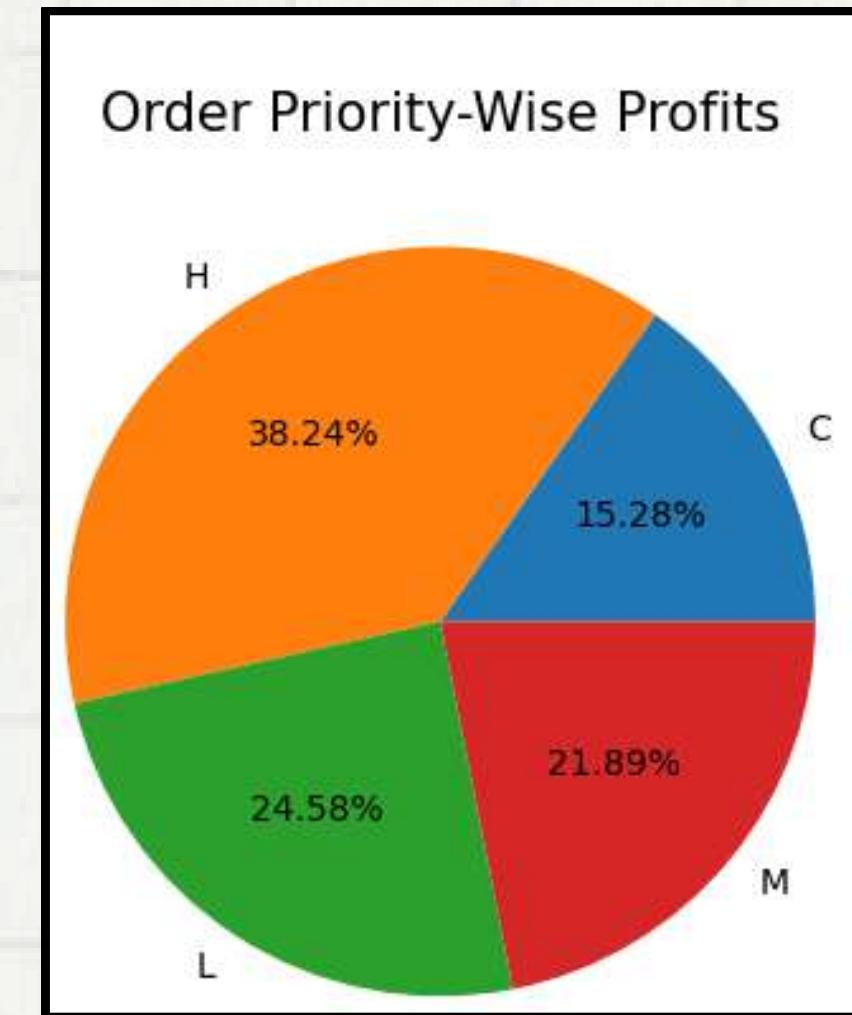
Take a look at the Channel-Wise profit pie chart which shows that 56.42% Profits were from Offline and 43.58% Profit were from Online.

**After Looking both of the Pie Chart you can see that ratio between Offline and Online is approximately same in both Sales and Profits which concludes that they have a great profit margin and doesn't take significant loss.**

# Order Priority-Wise Analysis



- C - Critical
- H - High
- M - Medium
- L - Low



This sales pie chart shows that 35.49% of the sales were from High Priority, 26.67% of the sales were from Low Priority, 24.11% of the sales were from Medium Priority and 13.73% of the sales were from Critical Priority. Highest was from High Priority and Lowest was from Critical Priority.

This profits pie chart shows that 38.24% of the profits were from High Priority, 24.58% of the profits were from Low Priority, 21.89% of the profits were from Medium Priority and 15.28% of the profits were from Critical Priority. The approximately similarity between sales and Profit graph show that all of them had a great profit margin and hadn't got any significant loss.

# Data Visualization of Bird Strikes

# What is Bird Strike?

A bird strike is strictly defined as a collision between a bird and an aircraft which is in flight or on a take-off or landing roll

## What is the goal of the project?

The goal of the project is to analyze the given dataset and answer the given case studies questions.

## Information about the Dataset.

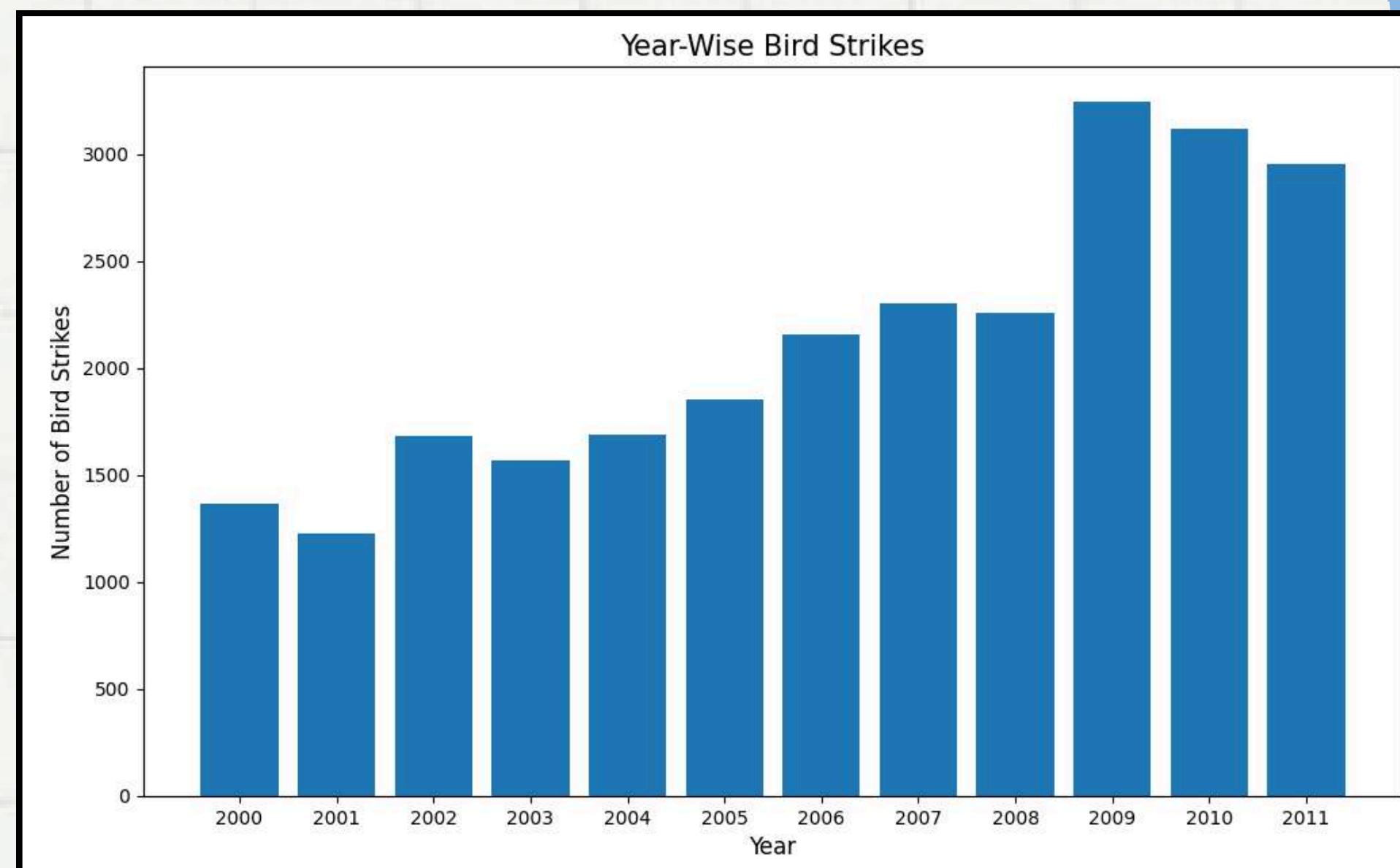
Given dataset contain the bird strike data from 2000 to 2011. Given dataset contain total of 26 columns. Given dataset contain many null values that was later fixed by me, to know what i did to fix those null values go to my code file of this project.



# Yearly Analysis of Bird Strike

Take a look at the bar graph here which is of number of bird strike incidents happened per year from 2000 to 2011. We can see that there is constant growth over year with some getting a slight decrement in the number of bird strike incidents. But you can see a constant growth in the graph.

You can see there was a very largest amount of bird strike incident happened in the year of 2009, 2010, 2011 with highest in 2009 later get a slight decrement in 2010 and 2011 but still it was very large amount.



# Top 10 US Airlines in terms of having encountered bird strikes

You can see here is the Top 10 US Airlines in term of encountered bird strikes. Highest number of bird strikes encountered by an US Airline was **Southwest Airlines(4628)** and the lowest by **US Airways(540)**.

Aircraft: Airline/Operator	count
SOUTHWEST AIRLINES	4628
BUSINESS	3074
AMERICAN AIRLINES	2058
DELTA AIR LINES	1349
AMERICAN EAGLE AIRLINES	932
SKYWEST AIRLINES	891
US AIRWAYS*	797
JETBLUE AIRWAYS	708
UPS AIRLINES	590
US AIRWAYS	540

# Top 50 Airports with most incidents of bird strikes

Airport: Name	count
DALLAS/FORT WORTH INTL ARPT	803
SACRAMENTO INTL	676
SALT LAKE CITY INTL	479
DENVER INTL AIRPORT	476
KANSAS CITY INTL	452
PHILADELPHIA INTL	442
ORLANDO INTL	408
BALTIMORE WASH INTL	401
LOUISVILLE INTL ARPT	395
JOHN F KENNEDY INTL	390
CHARLOTTE/DOUGLAS INTL ARPT	367
NASHVILLE INTL	364
LAMBERT-ST LOUIS INTL	363
CHICAGO O'HARE INTL ARPT	332
DETROIT METRO WAYNE COUNTY ARPT	321
PORTLAND INTL (OR)	313
NEWARK LIBERTY INTL ARPT	305
CINCINNATI/NORTHERN KENTUCKY INTL ARPT	302
ATLANTA INTL	296
CHICAGO MIDWAY INTL ARPT	296
HOUSTON-HOBBY	294
FORT LAUDERDALE/HOLLYWOOD INTL	277
WASHINGTON DULLES INTL ARPT	270
EPPLY AIRFIELD	269
LAGUARDIA NY	263
LOGAN INTL	255
DALLAS LOVE FIELD ARPT	253
METRO OAKLAND INTL	252
MINNEAPOLIS-ST PAUL INTL	251
GREATER PITTSBURGH	237
AUSTIN-BERGSTROM INTL	236
SAN FRANCISCO INTL ARPT	226
NEW ORLEANS INTL	219
LOS ANGELES INTL	217
SOUTHWEST FLORIDA INTL ARPT	215
CLEVELAND-HOPKINS INTL ARPT	214
BIRMINGHAM-SHUTTLESWORTH INTL	211
MIAMI INTL	202
MINETA SAN JOSE INTL	201
GEORGE BUSH INTERCONTINENTAL	199
PHOENIX SKY HARBOR	196
LIHUE ARPT	195
SAN ANTONIO INTL	194
BUFFALO-NIAGARA INTL	182
HONOLULU INTL ARPT	170
PORT COLUMBUS INTL	166
RALEIGH-DURHAM INTL	165
RONALD REAGAN WASHINGTON NATL	162
INDIANAPOLIS INTL	160
JACKSONVILLE INTL	155

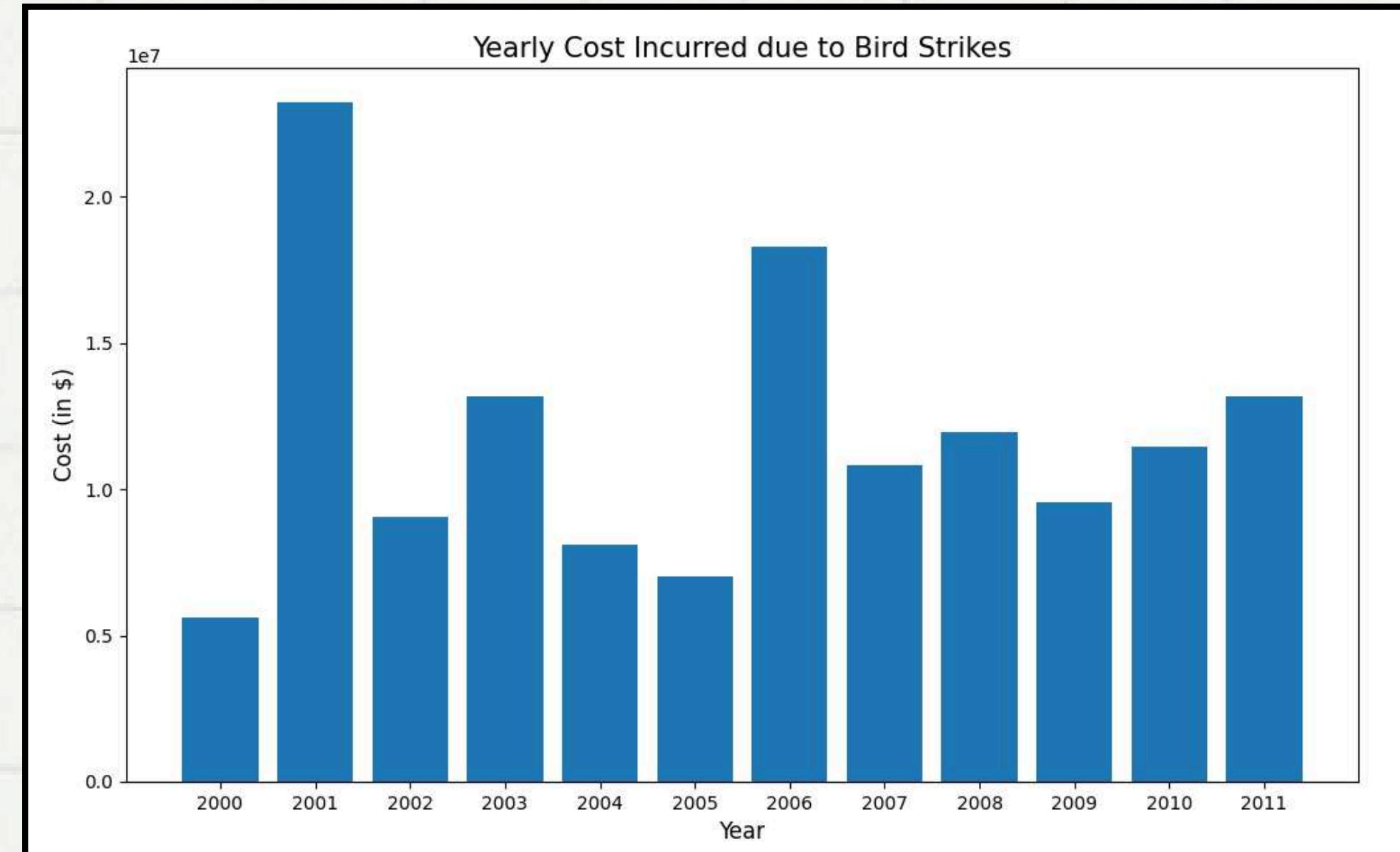
Above is the Top 50 Airport with most incidents of bird strikes which gives us a result **DALLAS/FORT WORTH INTL ARPT** has the highest incidents of bird strikes(803) and **JACKSONVILLE INTL ARPT** has the lowest incidents of bird strikes(155).

# Yearly Cost Incurred due to Bird Strikes

Now take a look yearly cost due to bird strikes graph that shows us that 2001 it was costed the most due to bird strikes and least was in 2000.

We can notice that earlier we found that there were highest incident of Bird Strike in 2009 to 2011, but still there is not much cost caused in those years. This means that they have developed many way to tackle bird strike we less or no cost.

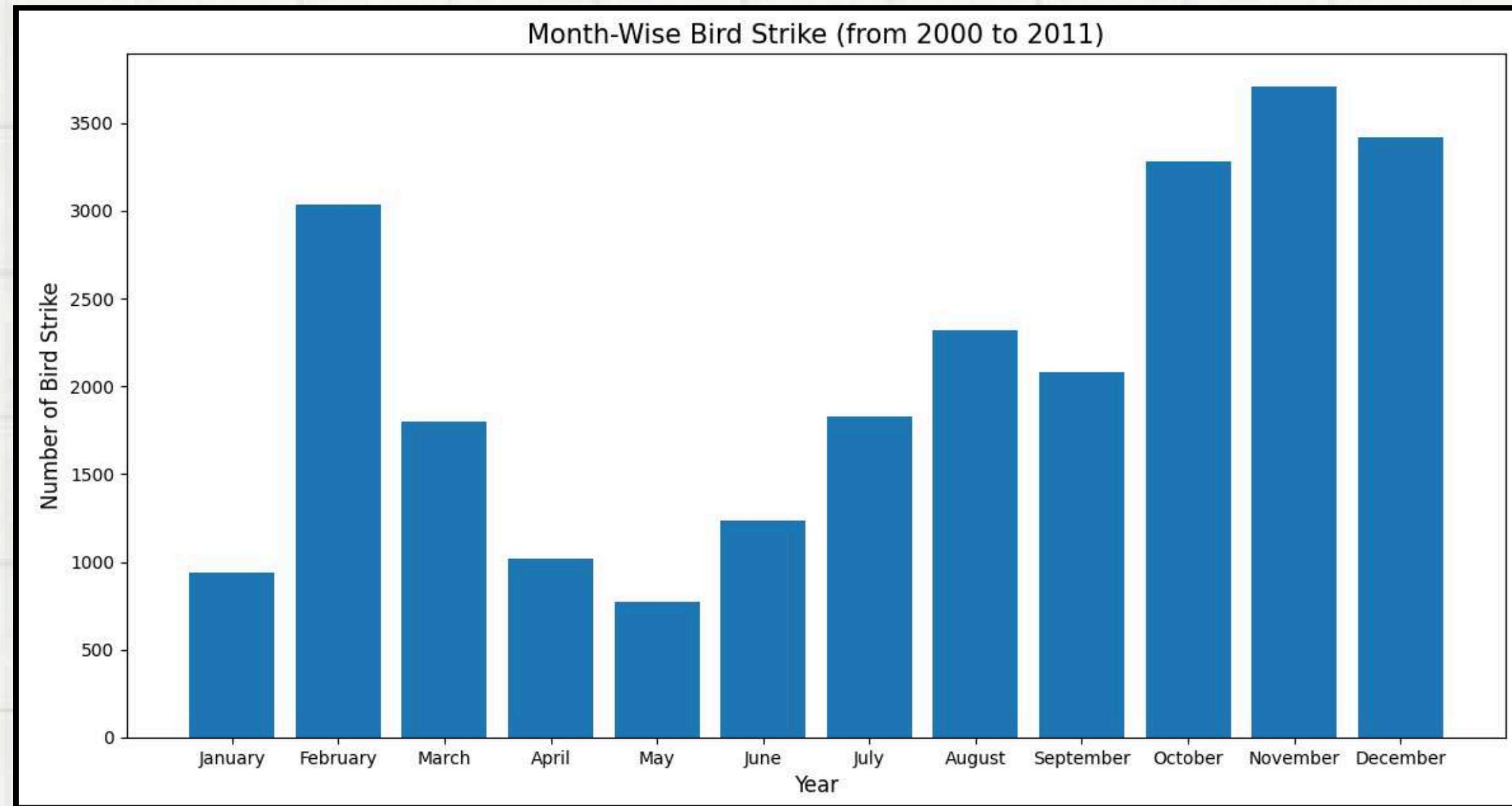
On the other hand we can see that there was highest cost in 2001 but there was not much Bird Strike occurred in that year. This means that at that time there was not that much way was developed to tackle Bird Strike at that time.



# When do most bird strikes occur?

We know that Bird Strike is a type of incident that can happen anytime but during migratory season the Bird Strike incident got a major increment :

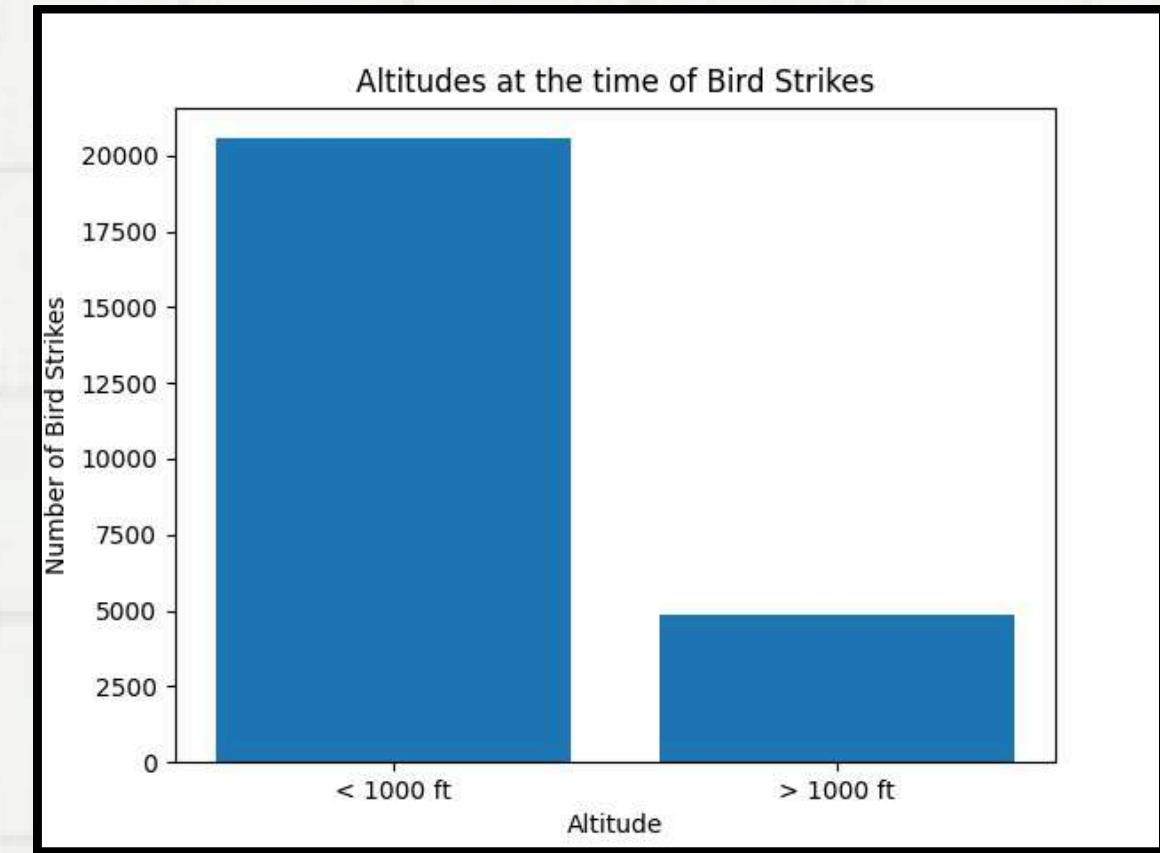
- **Spring Migration** – This Migration period is from Late-February to Early-June. Birds tend to migrate faster and more urgently to reach their breeding grounds. Birds may fly at different altitudes to take advantage of favorable weather conditions, which can sometimes overlap with aircraft flight paths, especially during lower-altitude flight phases like takeoff and landing.
- **Fall Migration** – The fall migration period is often more prolonged, with birds migrating over a longer span of time. The presence of a higher number of juvenile birds, which are less experienced in avoiding aircraft. Many bird species form larger flocks during fall migration, which can increase the likelihood of encountering multiple bird strikes.



# Altitude of aeroplanes at the time of strike

Now take a look at the Altitude of the aeroplanes at the time of strike this shows that over 80% times Bird Strike occur when aircraft was below or equal to 1000ft.

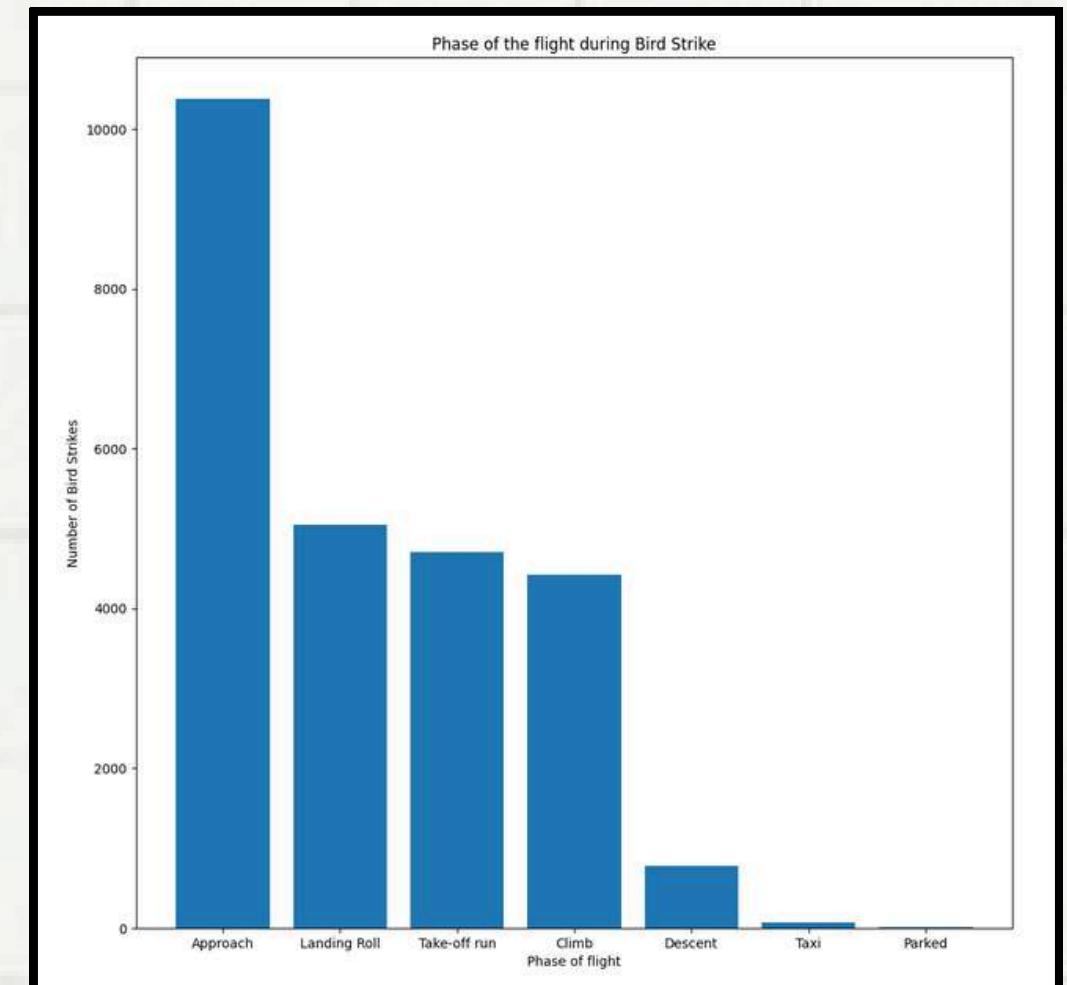
In which over 20000 bird strikes incident occurs when aeroplane was below 1000ft and close to 5000 bird strikes incident occurs when aeroplane was above 1000ft.



## Phase of flight at the time of the strike

- **Parked** – When Aircraft is parked at the airport.
- **Taxi** – Movement of Aircraft on the ground towards the runway.
- **Take-off run** – The phase where the aircraft accelerates along the runway until it gains enough lift to air.
- **Climb** – The phase where the aircraft continues to ascend to its cruising altitude.
- **Descent** – The phase where the aircraft decreases altitude in preparation for landing.
- **Approach** – The phase where the aircraft aligns with the runway and prepares for landing.
- **Landing roll** – The phase where the aircraft touches down on the runway and slows to taxiing speed.

After seeing the graph we can see that most amount of bird strikes occurs when aircraft is in **Approach** phase and the least in **Parked** phase.



# Average Altitude of the aeroplanes in different phases at the time of strike

We can get this average altitude of the aeroplane in different phases by grouping Phase of flight with altitude mean.

We got the highest average altitude in Descent phase of flight and zero or almost zero average altitude in Landing Roll, Parked, Take-off run and Taxi. Got approximately 1000ft in Approach and Climb.

When: Phase of flight	Feet above ground
Approach	1001.733385
Climb	1199.654324
Descent	5933.960052
Landing Roll	0.000000
Parked	0.000000
Take-off run	0.101040
Taxi	0.000000

# Effect of Bird Strikes & Impact on Flight

You can see that the highest number of bird strikes incident gives an **Unknown** Impact to the flight and the lowest number of bird strike incidents give **Engine Shut Down** Impact to the flight.

Note – Here **Unknown** is the value we replaces with null values

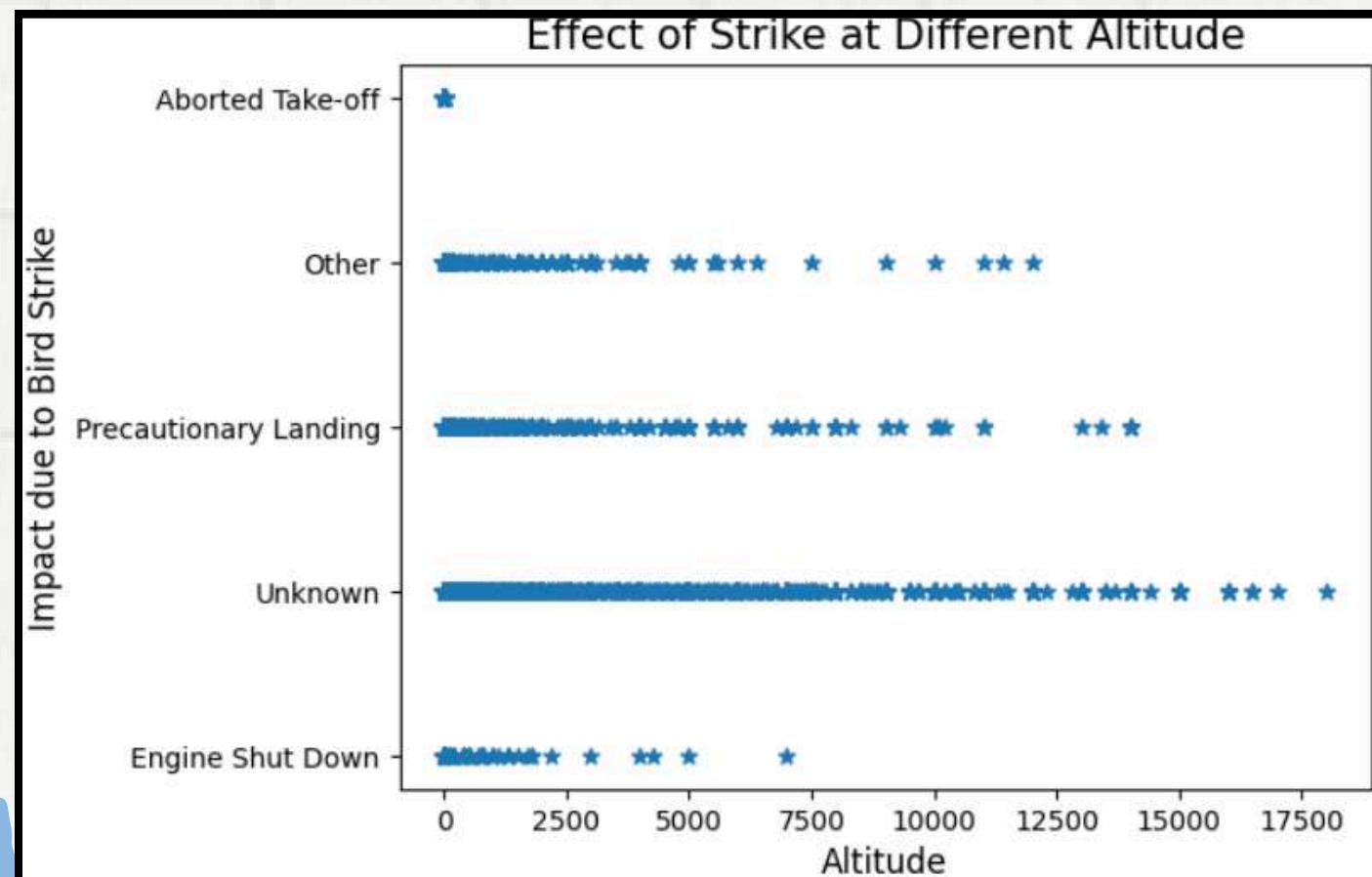
Effect: Impact to flight	count
Unknown	23351
Precautionary Landing	1121
Aborted Take-off	479
Other	390
Engine Shut Down	88

## Effect of Strike at Different Altitude

You can see the scatterplot of the Impact due to bird strike and Altitude of the aircraft which shows the effect of strike at different altitude.

- **Aborted Take-off** – All are at zero altitude.
- **Other** – Mostly are in range 0 to 3000 approx.
- **Precautionary Landing** – Mostly are in range 0 to 5000 approx.
- **Unknown** – Mostly are in range 0 to 12500 approx.
- **Engine Shut Down** – Mostly are in range 0 to 2000 approx.

Note – Here **Unknown** is the value we replaces with null values



# Were Pilots Informed?

We get 42.71% of times pilot was warned about the upcoming bird strike and 57.29% of times pilot was not warned about the upcoming bird strike.

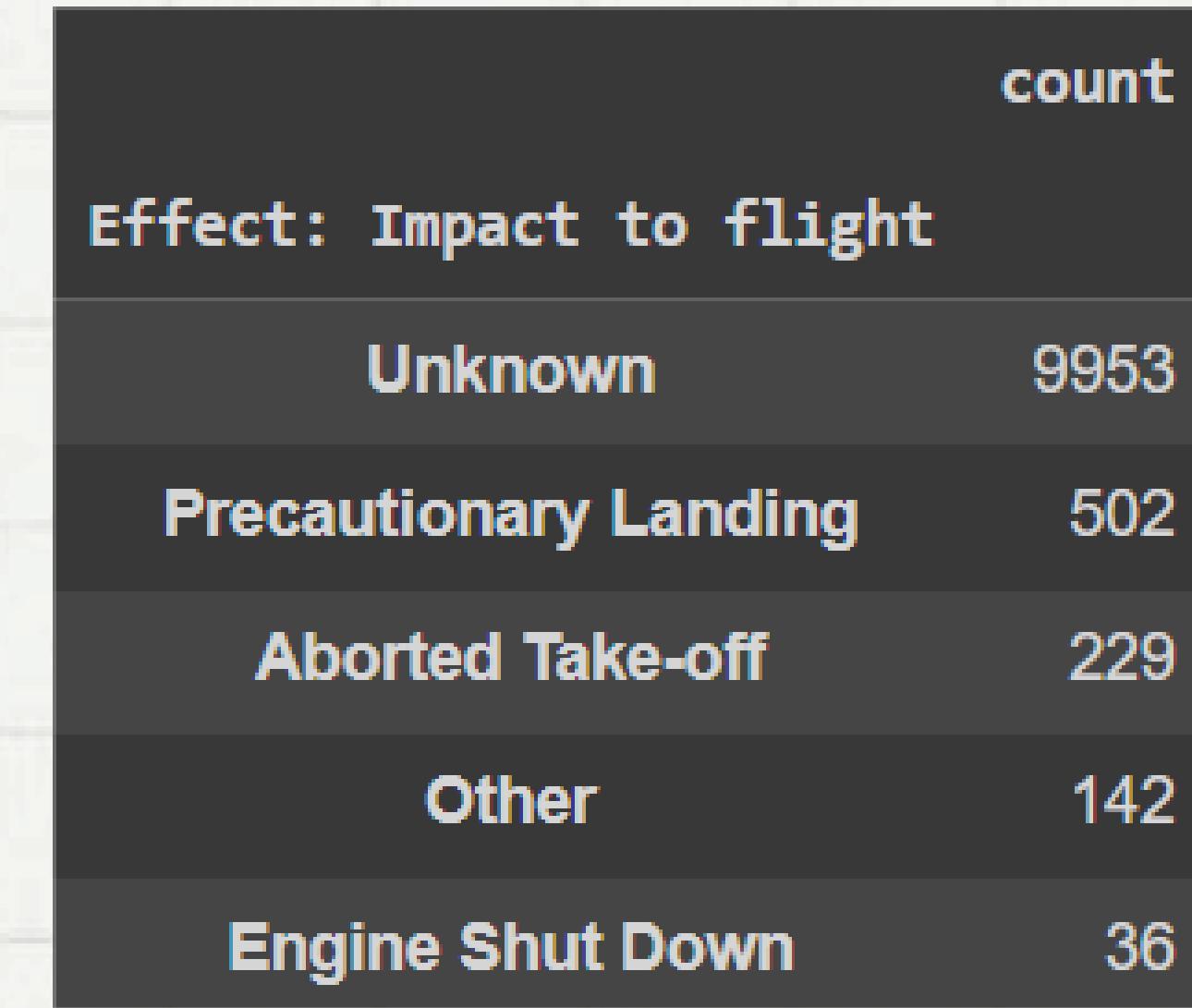
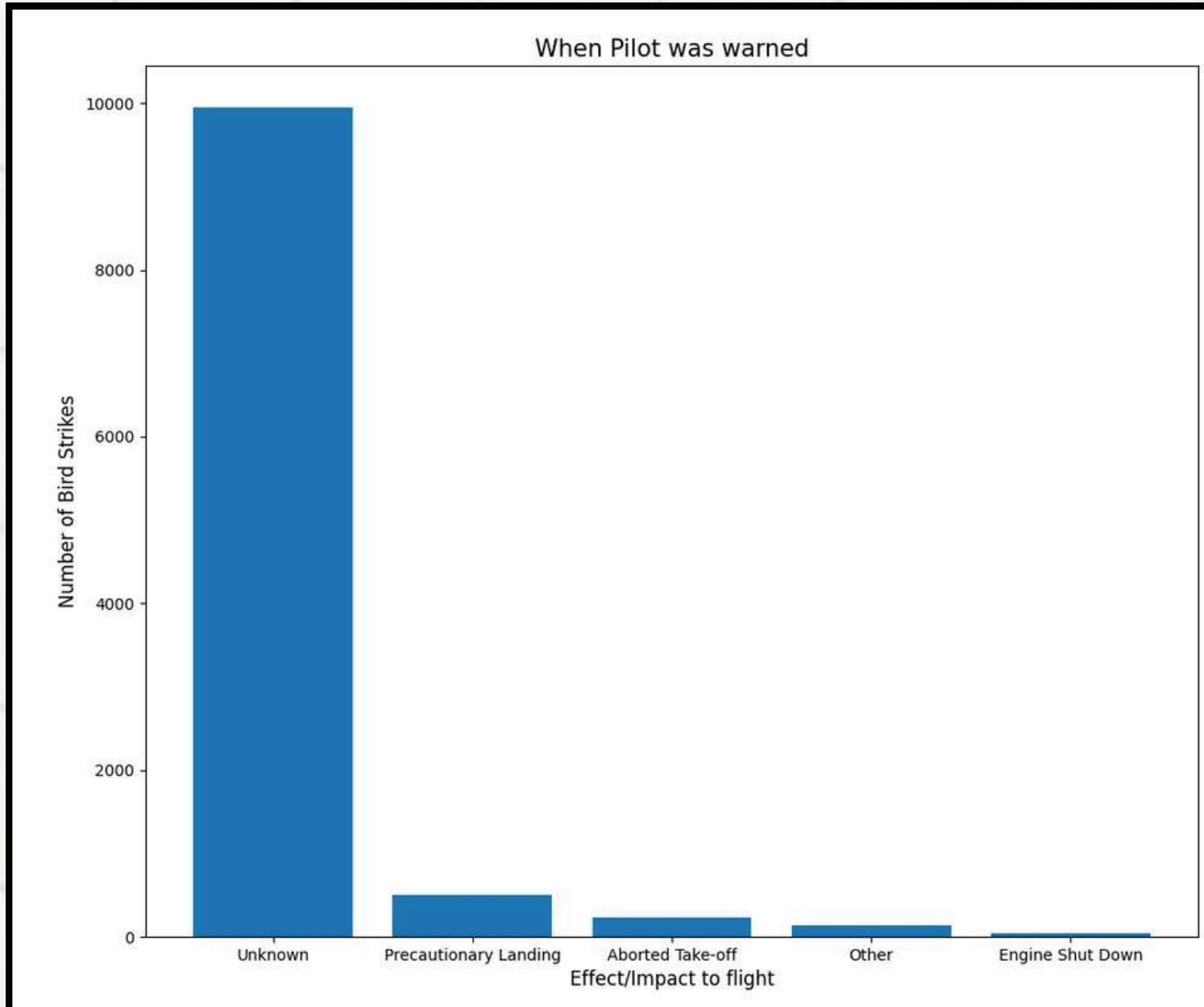
Pilot warned of birds or wildlife?	count
N	14567
Y	10862

# Damage Taken?

We get 9.65% of times bird strike caused damage to aircraft and 90.35% of times bird strike caused no damage to aircraft.

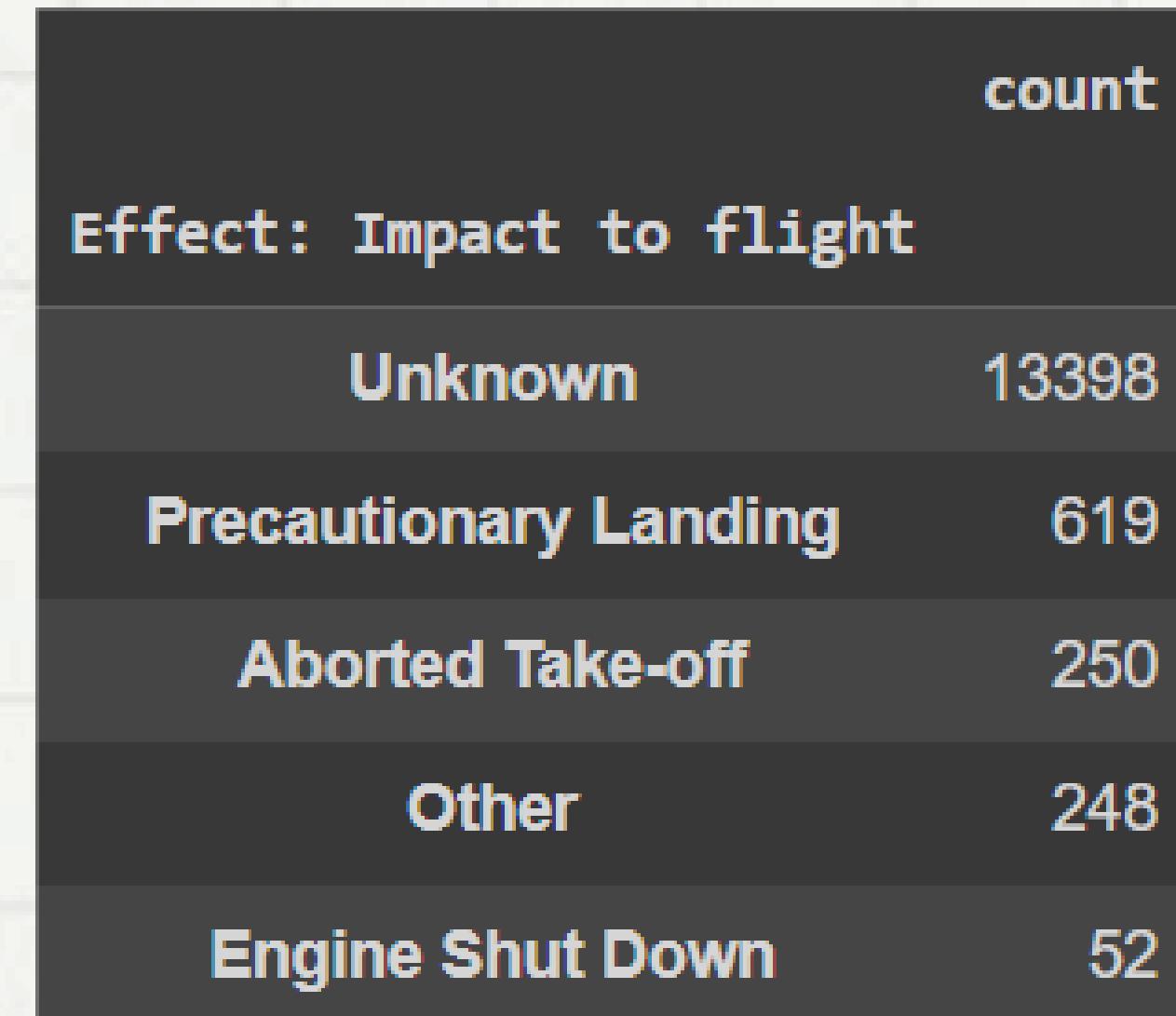
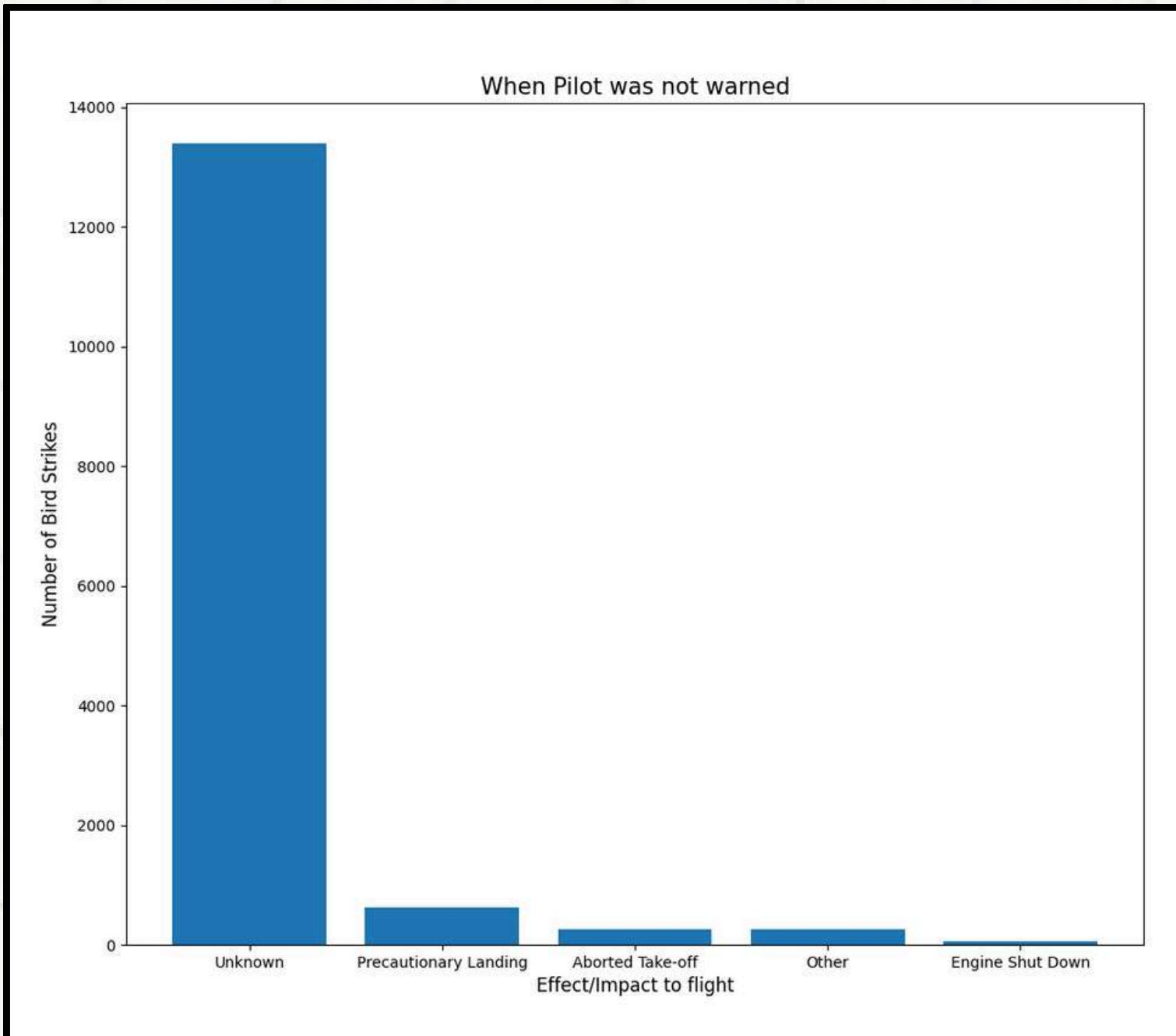
Effect: Indicated Damage	count
No damage	22975
Caused damage	2454

# When Pilot was warned



Note – Here **Unknown** is the value we replaces with null values

# When Pilot was not warned



Note - Here Unknown is the value we replaces with null values

# Range of the number of bird that got struck

	count
Wildlife: Number struck	
1	20790
2 to 10	4319
11 to 100	312
Over 100	8

We can say that if there is a Bird Strike then there is-

- 1.82% chances that there is only 1 bird.
- 2.16.8% chances that there is 2 to 10 birds.
- 3.1% chances that there 11 to 100 birds.
- 4.0.2% chances that there over 100 birds.

## Was the Aircraft Large or not?

	count
Is Aircraft Large?	
No	17027
Yes	8402

When Bird Strike happen then there is -

- 1.67% chances that the Aircraft is Not Large.
- 2.33% chances that the Aircraft is Large.

## Number of people getting injured

	count
Number of people injured	
0	25416
1	9
2	3
6	1

We can see that there are very less probability of people getting injured due to Bird Strike on the Aircraft.

# The End

<https://github.com/ARNAVKS>