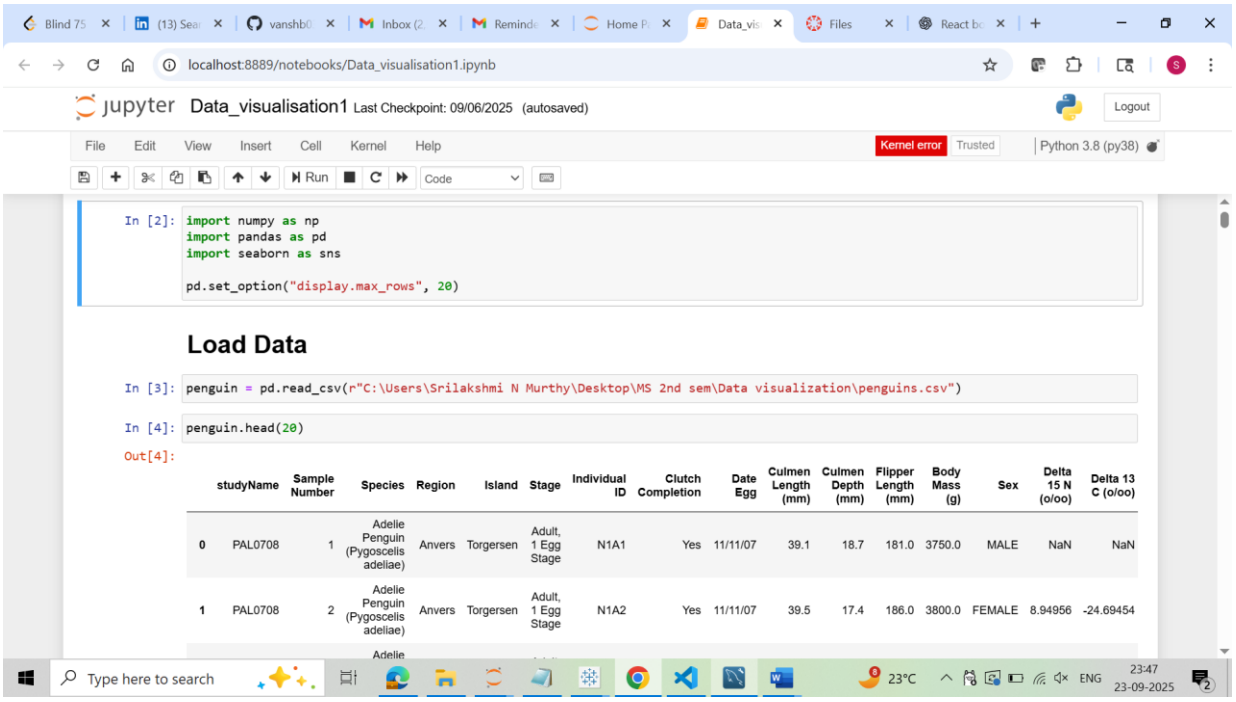
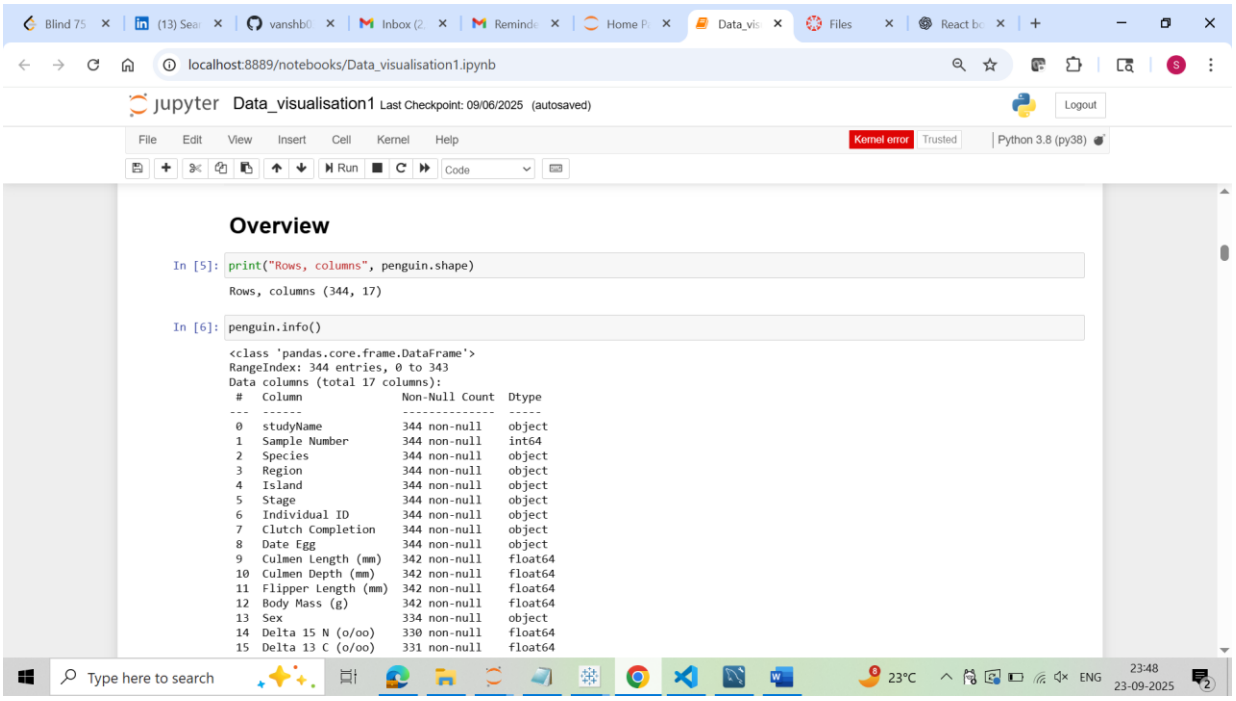
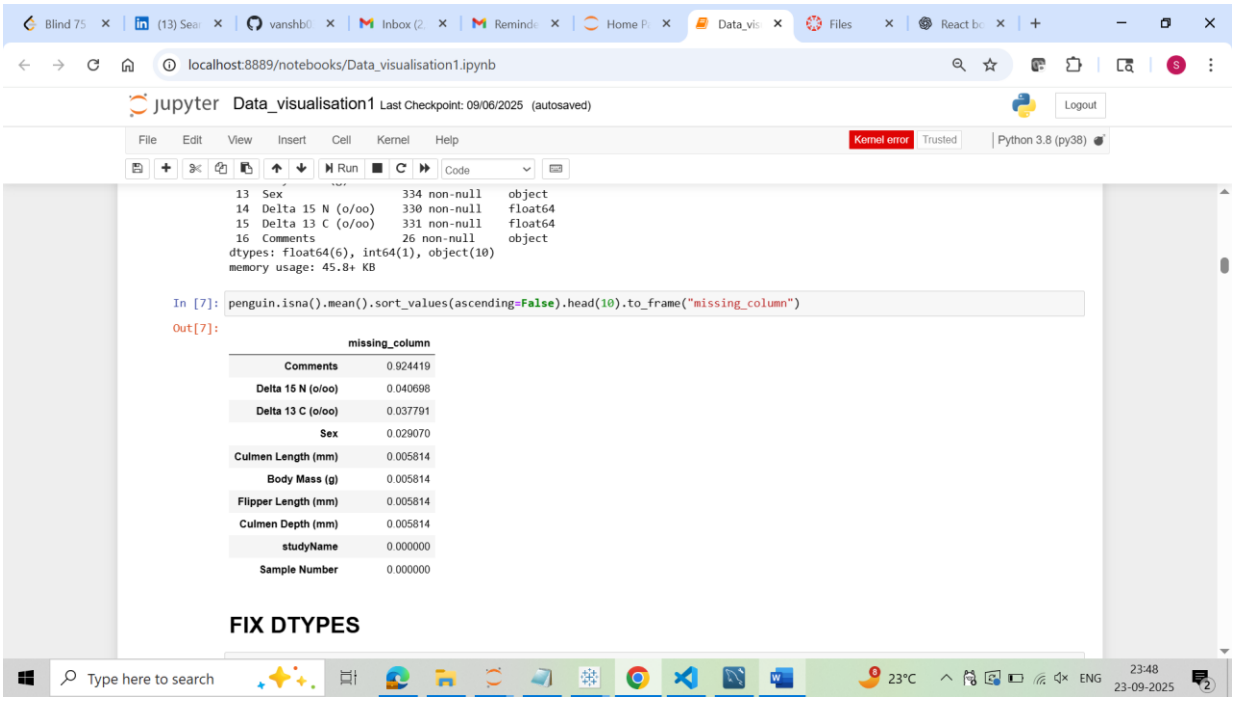
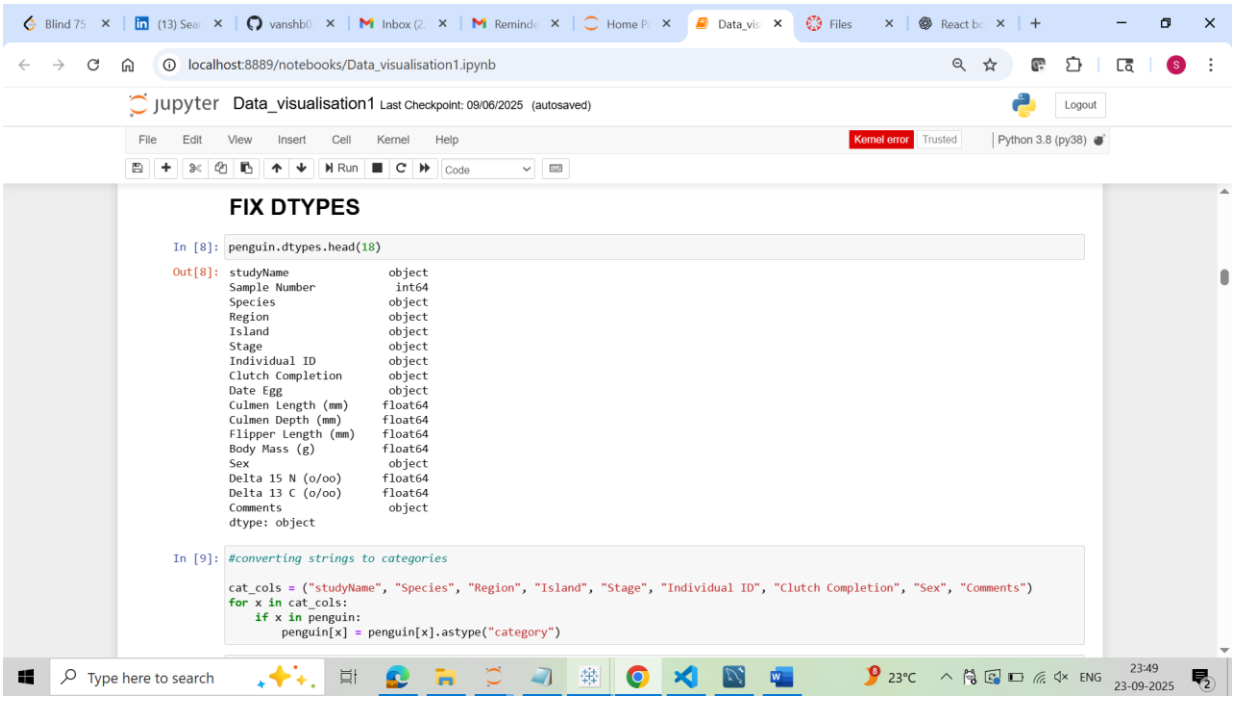
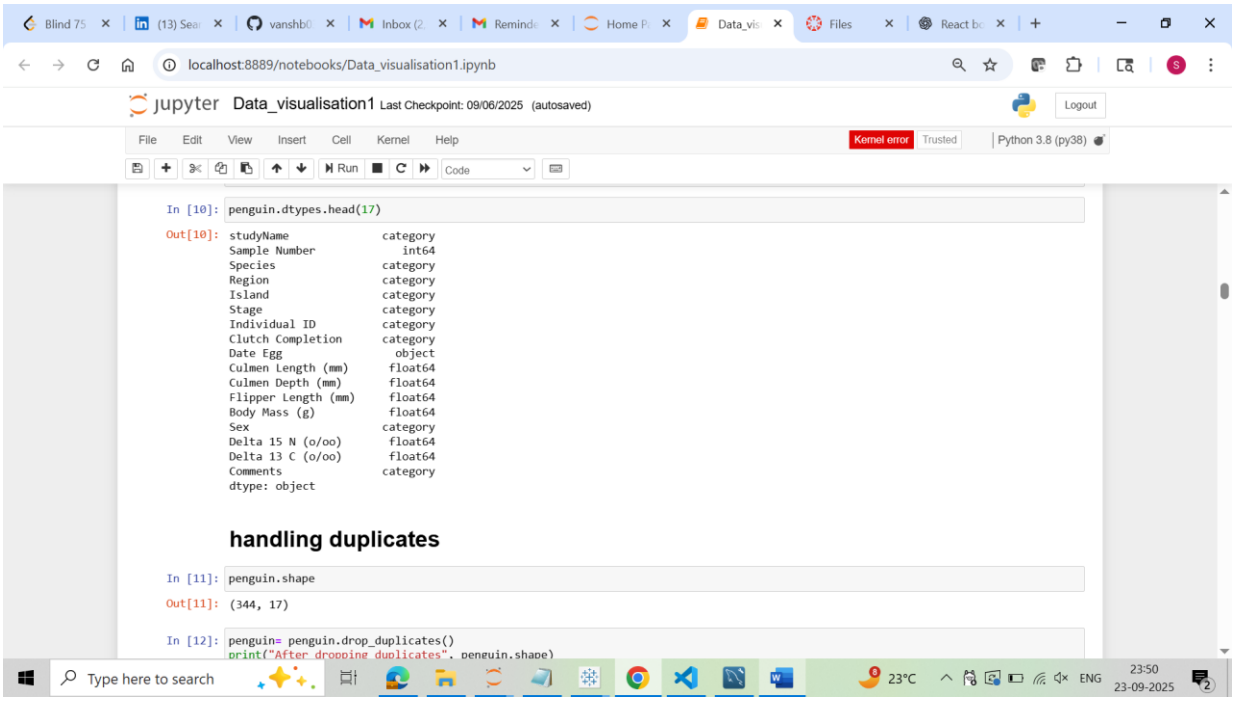
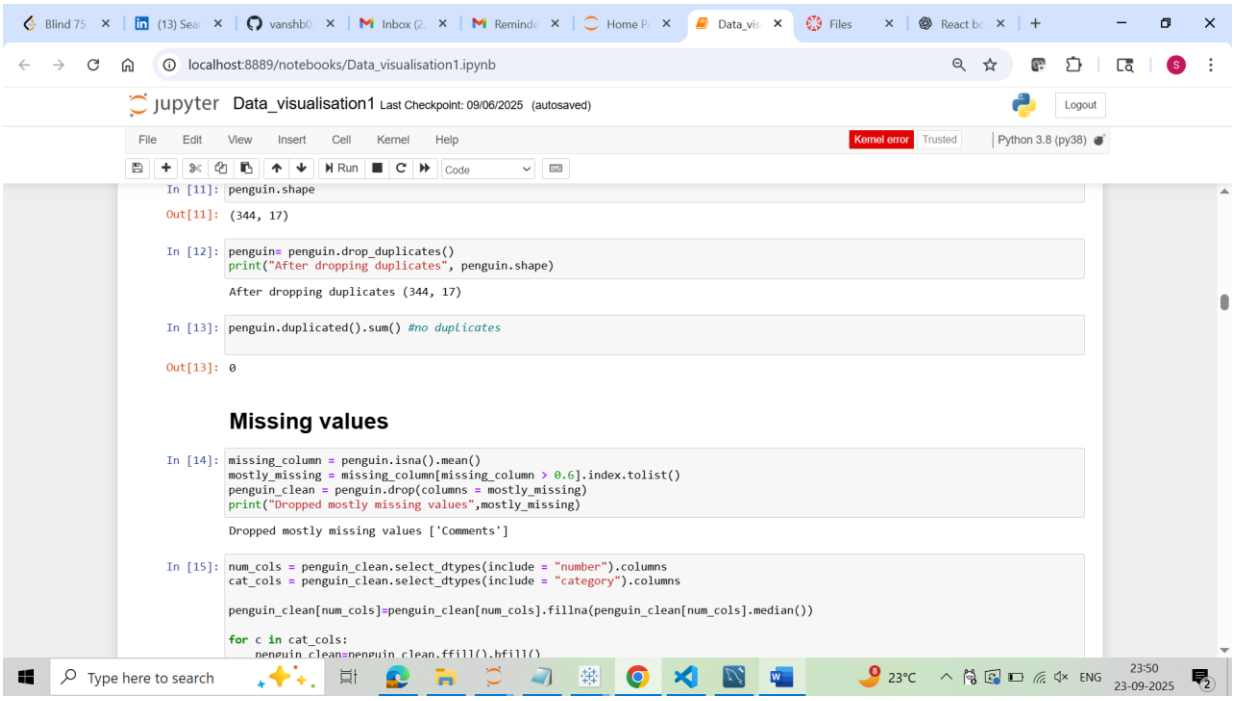

localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error Trusted Python 3.8 (py38)
Run Code
In [2]:
import numpy as np
import pandas as pd
import seaborn as sns
pd.set_option("display.max_rows", 20)
Load Data
In [3]:
penguin = pd.read_csv(r"C:\Users\Srilakshmi N Murthy\Desktop\MS 2nd sem\Data visualization\penguins.csv")
In [4]:
penguin.head(20)
Out[4]:
studyName Sample Number Species Region Island Stage Individual ID Clutch Completion Date Egg Culmen Length (mm) Culmen Depth (mm) Flipper Length (mm) Body Mass (g) Sex Delta 15 N (o/oo) Delta 13 C (o/oo)
0 PAL0708 1 Adelie Penguin (Pygoscelis adeliae) Anvers Torgersen Adult, 1 Egg Stage N1A1 Yes 11/11/07 39.1 18.7 181.0 3750.0 MALE NaN NaN
1 PAL0708 2 Adelie Penguin (Pygoscelis adeliae) Anvers Torgersen Adult, 1 Egg Stage N1A2 Yes 11/11/07 39.5 17.4 186.0 3800.0 FEMALE 8.94956 -24.69454
Type here to search
23°C
ENG
23:47
23-09-2025

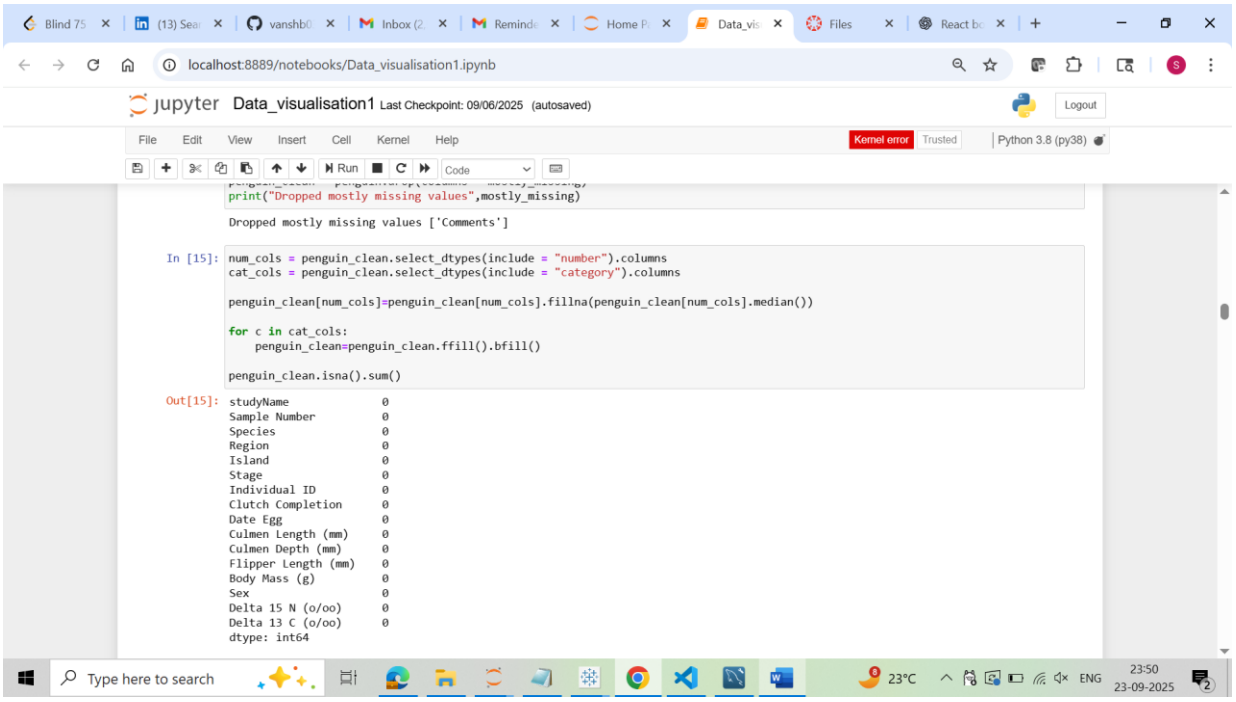

localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error Trusted Python 3.8 (py38)
Run Code
Overview
In [5]:
print("Rows, columns", penguin.shape)
Rows, columns (344, 17)
In [6]:
penguin.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 344 entries, 0 to 343
Data columns (total 17 columns):
Column Non-Null Count Dtype
0 studyName 344 non-null object
1 Sample Number 344 non-null int64
2 Species 344 non-null object
3 Region 344 non-null object
4 Island 344 non-null object
5 Stage 344 non-null object
6 Individual ID 344 non-null object
7 Clutch Completion 344 non-null object
8 Date Egg 344 non-null object
9 Culmen Length (mm) 342 non-null float64
10 Culmen Depth (mm) 342 non-null float64
11 Flipper Length (mm) 342 non-null float64
12 Body Mass (g) 342 non-null float64
13 Sex 334 non-null object
14 Delta 15 N (o/oo) 330 non-null float64
15 Delta 13 C (o/oo) 331 non-null float64
Type here to search
23°C
ENG
23:48
23-09-2025

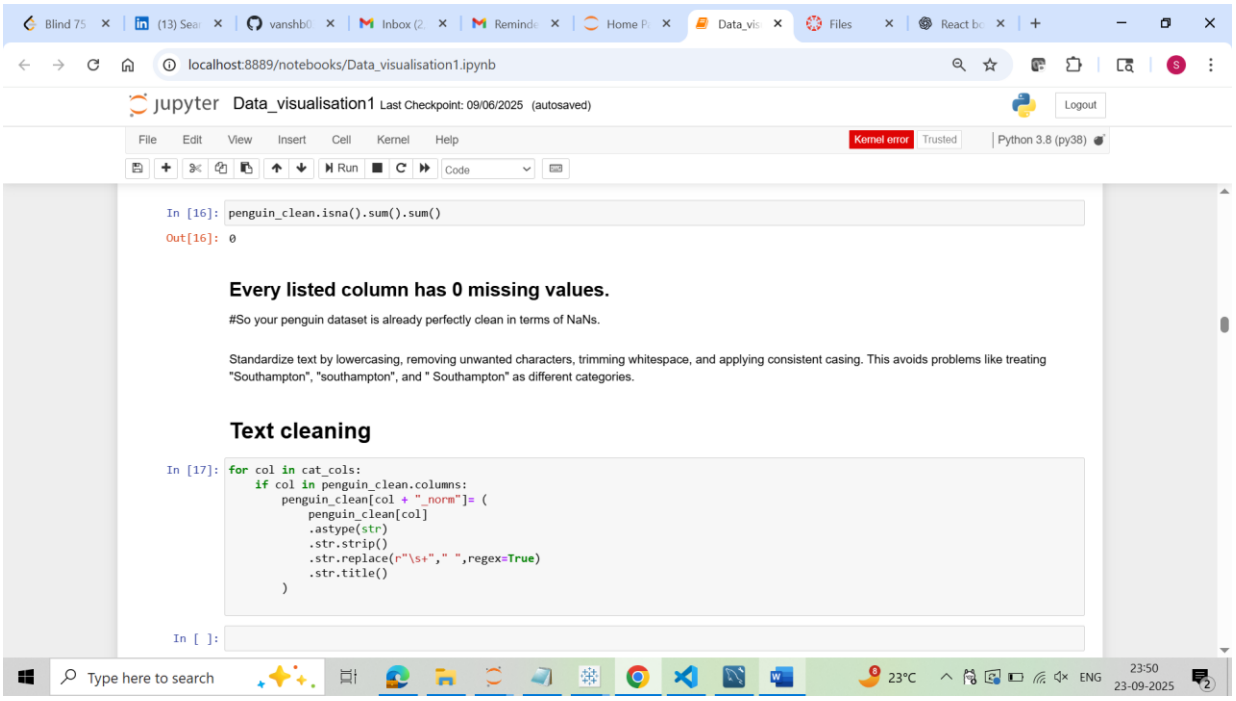

localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error Trusted Python 3.8 (py38)
Run Code
13 Sex 334 non-null object
14 Delta 15 N (o/oo) 330 non-null float64
15 Delta 13 C (o/oo) 331 non-null float64
16 Comments 26 non-null object
dtypes: float64(6), int64(1), object(10)
memory usage: 45.8+ KB
In [7]: penguin.isna().mean().sort_values(ascending=False).head(10).to_frame("missing_column")
Out[7]:
missing_column
Comments 0.924419
Delta 15 N (o/oo) 0.040698
Delta 13 C (o/oo) 0.037791
Sex 0.029070
Culmen Length (mm) 0.005814
Body Mass (g) 0.005814
Flipper Length (mm) 0.005814
Culmen Depth (mm) 0.005814
studyName 0.000000
Sample Number 0.000000
FIX DTYPES
Type here to search
23°C
ENG
23:48
23-09-2025

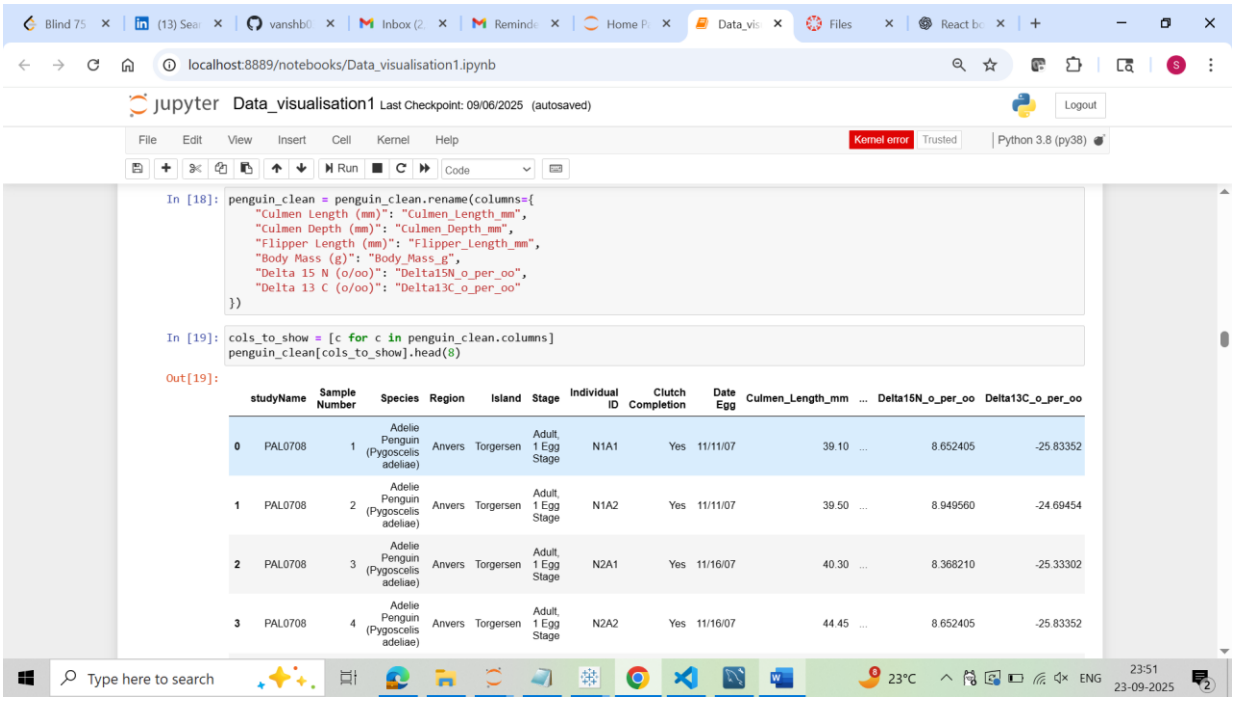

localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error Trusted Python 3.8 (py38)
Run Code
FIX DTYPES
In [8]: penguin.dtypes.head(18)
Out[8]:
studyName object
Sample Number int64
Species object
Region object
Island object
Stage object
Individual ID object
Clutch Completion object
Date Egg object
Culmen Length (mm) float64
Culmen Depth (mm) float64
Flipper Length (mm) float64
Body Mass (g) float64
Sex object
Delta 15 N (o/oo) float64
Delta 13 C (o/oo) float64
Comments object
dtype: object
In [9]:
#converting strings to categories
cat_cols = ("studyName", "Species", "Region", "Island", "Stage", "Individual ID", "Clutch Completion", "Sex", "Comments")
for x in cat_cols:
if x in penguin:
penguin[x] = penguin[x].astype("category")
Type here to search
23°C
ENG
23:49
23-09-2025

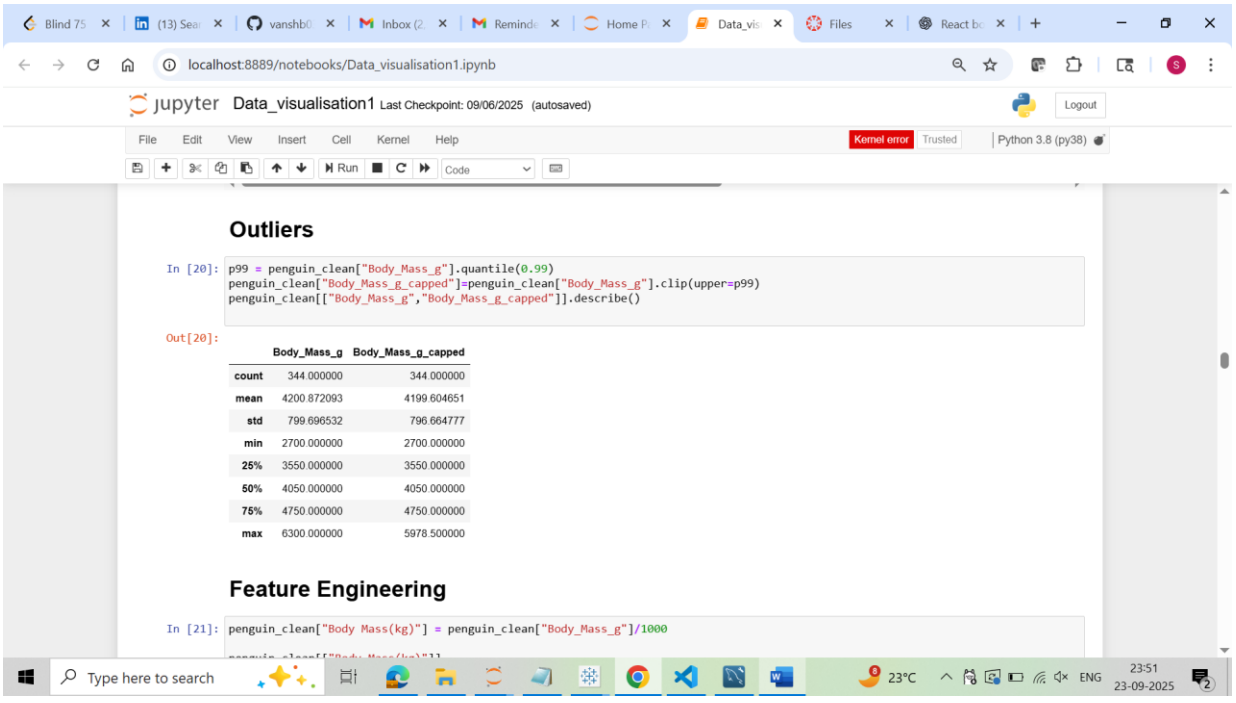

localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error Trusted Python 3.8 (py38)
Run Code
In [10]: penguin.dtypes.head(17)
Out[10]:
studyName category
Sample Number int64
Species category
Region category
Island category
Stage category
Individual ID category
Clutch Completion category
Date Egg object
Culmen Length (mm) float64
Culmen Depth (mm) float64
Flipper Length (mm) float64
Body Mass (g) float64
Sex category
Delta 15 N (o/oo) float64
Delta 13 C (o/oo) float64
Comments category
dtype: object
handling duplicates
In [11]: penguin.shape
Out[11]: (344, 17)
In [12]: penguin= penguin.drop_duplicates()
print("After dropping duplicates", penguin.shape)
Type here to search
23°C
ENG
23:50
23-09-2025

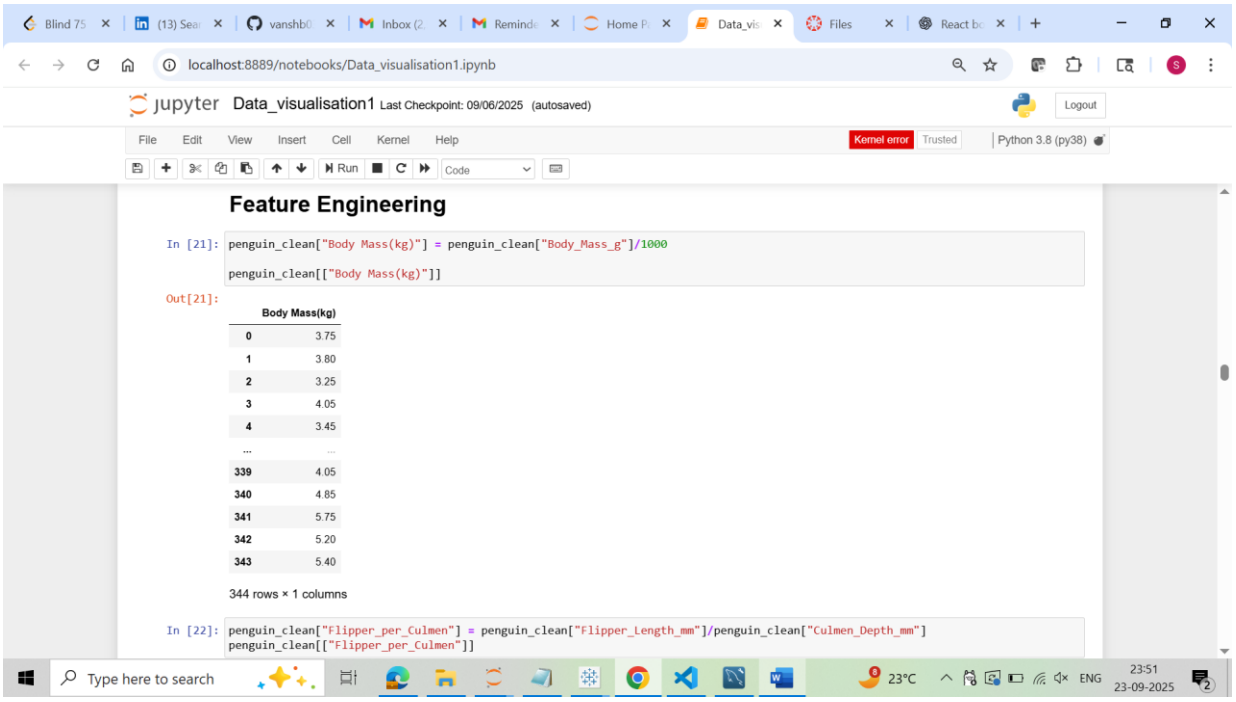

localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error Trusted Python 3.8 (py38)
Run Code
In [11]: penguin.shape
Out[11]: (344, 17)
In [12]: penguin= penguin.drop_duplicates()
print("After dropping duplicates", penguin.shape)
After dropping duplicates (344, 17)
In [13]: penguin.duplicated().sum() #no duplicates
Out[13]: 0
Missing values
In [14]: missing_column = penguin.isna().mean()
mostly_missing = missing_column[missing_column > 0.6].index.tolist()
penguin_clean = penguin.drop(columns = mostly_missing)
print("Dropped mostly missing values",mostly_missing)
Dropped mostly missing values ['Comments']
In [15]: num_cols = penguin_clean.select_dtypes(include = "number").columns
cat_cols = penguin_clean.select_dtypes(include = "category").columns
penguin_clean[num_cols]=penguin_clean[num_cols].fillna(penguin_clean[num_cols].median())
for c in cat_cols:
Type here to search
23°C
ENG
23:50
23-09-2025

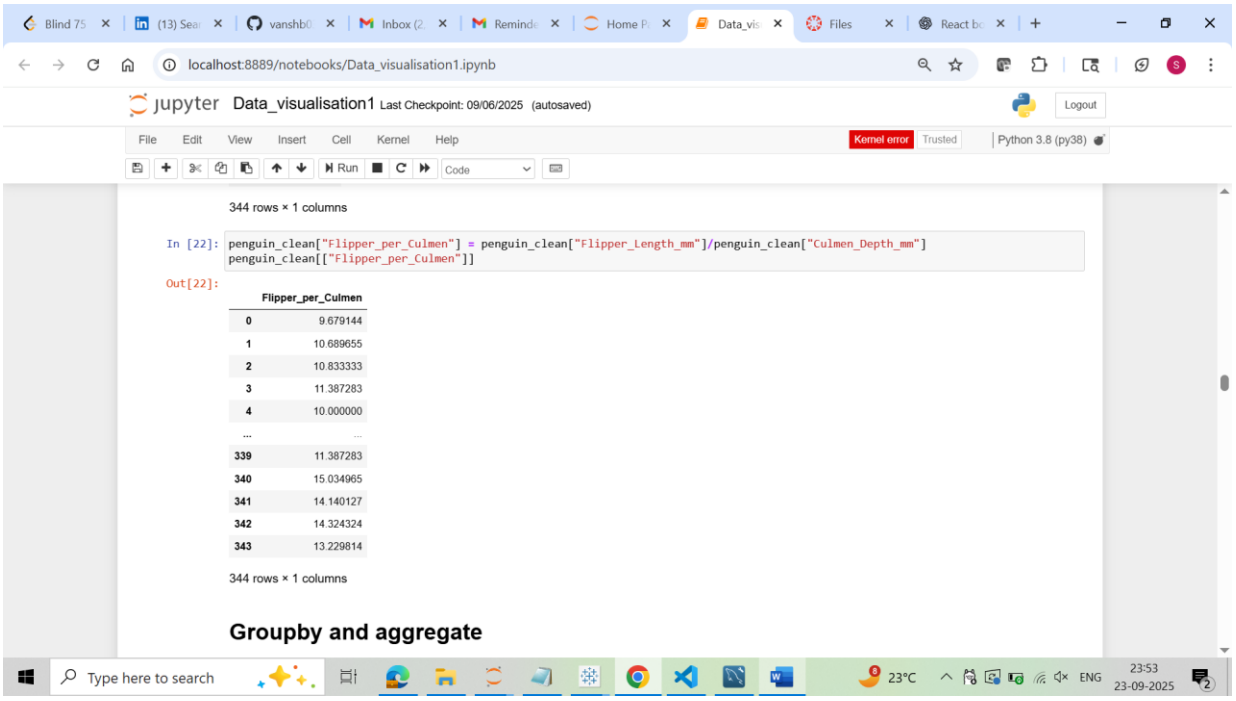

Blind 75
(13) Sea
vanshb0
Inbox (2
Remind
Home P
Data_vis
Files
React bc
localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error
Trusted
Python 3.8 (py38)
Run
Code
print("Dropped mostly missing values",mostly_missing)
Dropped mostly missing values ['Comments']
In [15]:
num_cols = penguin_clean.select_dtypes(include = "number").columns
cat_cols = penguin_clean.select_dtypes(include = "category").columns
penguin_clean[num_cols]=penguin_clean[num_cols].fillna(penguin_clean[num_cols].median())
for c in cat_cols:
penguin_clean=penguin_clean.ffill().bfill()
penguin_clean.isna().sum()
Out[15]:
studyName 0
Sample Number 0
Species 0
Region 0
Island 0
Stage 0
Individual ID 0
Clutch Completion 0
Date Egg 0
Culmen Length (mm) 0
Culmen Depth (mm) 0
Flipper Length (mm) 0
Body Mass (g) 0
Sex 0
Delta 15 N (o/oo) 0
Delta 13 C (o/oo) 0
dtype: int64
Type here to search
23°C
ENG
23:50
23-09-2025

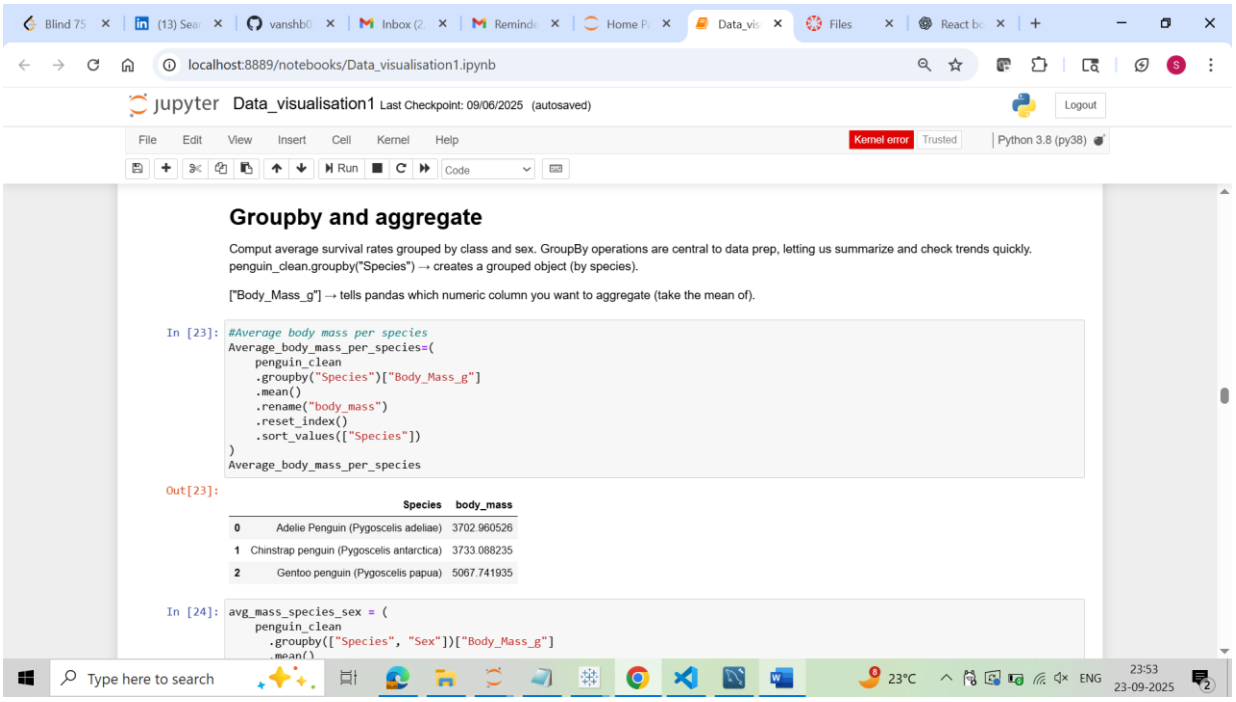

Blind 75
(13) Sea
vanshb0
Inbox (2
Remind
Home P
Data_vis
Files
React bc
localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error
Trusted
Python 3.8 (py38)
Run
Code
In [16]:
penguin_clean.isna().sum().sum()
Out[16]:
0
Every listed column has 0 missing values.
#So your penguin dataset is already perfectly clean in terms of NaNs.
Standardize text by lowercasing, removing unwanted characters, trimming whitespace, and applying consistent casing. This avoids problems like treating "Southampton", "southampton", and " Southampton" as different categories.
Text cleaning
In [17]:
for col in cat_cols:
if col in penguin_clean.columns:
penguin_clean[col + "_norm"]= (
penguin_clean[col]
.astype(str)
.str.strip()
.str.replace(r"\s+"," ",regex=True)
.str.title()
)
In [ ]:
Type here to search
23°C
ENG
23:50
23-09-2025

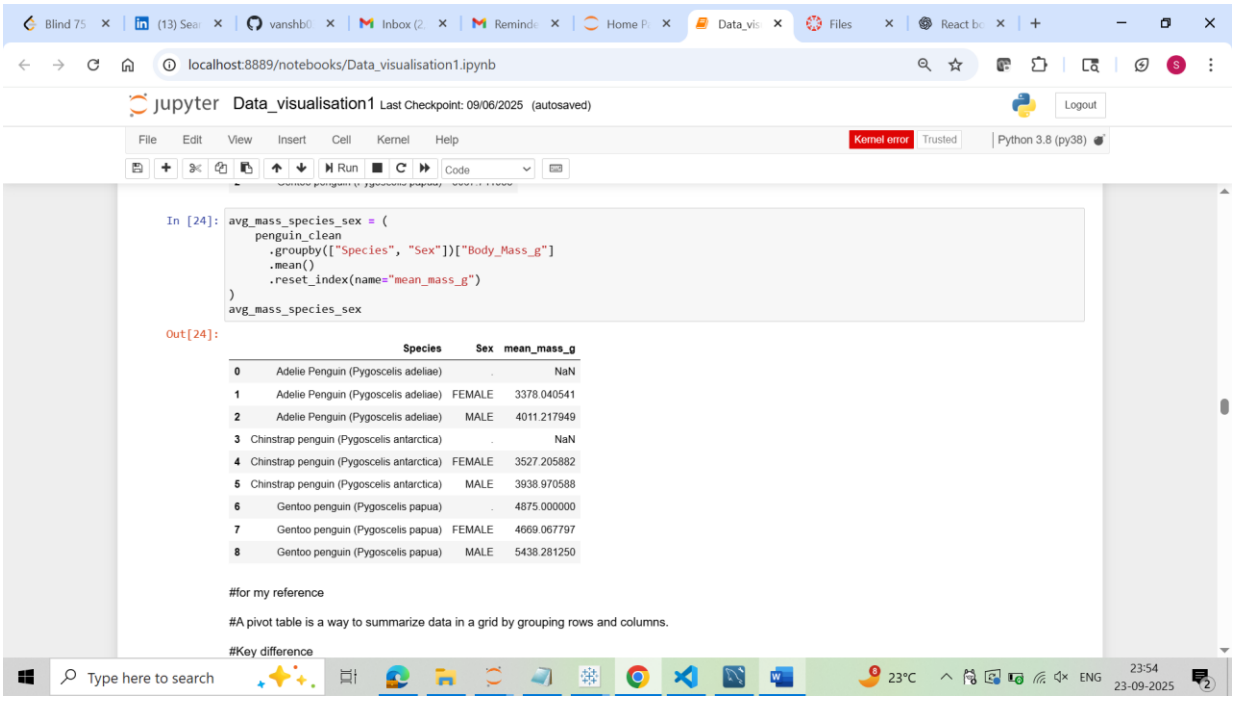

localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error Trusted Python 3.8 (py38)
Run Code
In [18]:
penguin_clean = penguin_clean.rename(columns={
"Culmen Length (mm)": "Culmen_Length_mm",
"Culmen Depth (mm)": "Culmen_Depth_mm",
"Flipper Length (mm)": "Flipper_Length_mm",
"Body Mass (g)": "Body_Mass_g",
"Delta 15 N (o/oo)": "Delta15N_o_per_oo",
"Delta 13 C (o/oo)": "Delta13C_o_per_oo"
})
In [19]:
cols_to_show = [c for c in penguin_clean.columns]
penguin_clean[cols_to_show].head(8)
Out[19]:
studyName Sample Number Species Region Island Stage Individual ID Clutch Completion Date Egg Culmen_Length_mm ... Delta15N_o_per_oo Delta13C_o_per_oo
0 PAL0708 1 Adelie Penguin (Pygoscelis adeliae) Anvers Torgersen Adult, 1 Egg Stage N1A1 Yes 11/11/07 39.10 ... 8.652405 -25.83352
1 PAL0708 2 Adelie Penguin (Pygoscelis adeliae) Anvers Torgersen Adult, 1 Egg Stage N1A2 Yes 11/11/07 39.50 ... 8.949560 -24.69454
2 PAL0708 3 Adelie Penguin (Pygoscelis adeliae) Anvers Torgersen Adult, 1 Egg Stage N2A1 Yes 11/16/07 40.30 ... 8.368210 -25.33302
3 PAL0708 4 Adelie Penguin (Pygoscelis adeliae) Anvers Torgersen Adult, 1 Egg Stage N2A2 Yes 11/16/07 44.45 ... 8.652405 -25.83352
Type here to search
23°C
ENG
23:51
23-09-2025

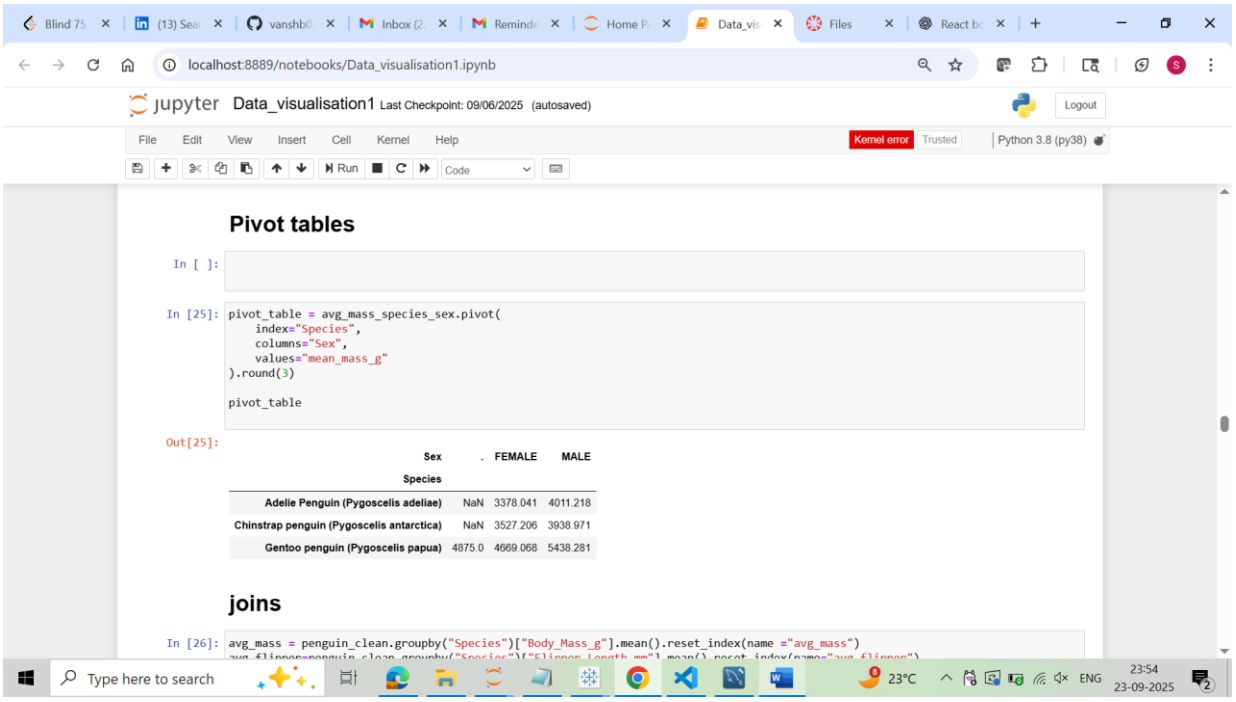

localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error Trusted Python 3.8 (py38)
Run Code
Outliers
In [20]:
p99 = penguin_clean["Body_Mass_g"].quantile(0.99)
penguin_clean["Body_Mass_g_capped"]=penguin_clean["Body_Mass_g"].clip(upper=p99)
penguin_clean[["Body_Mass_g","Body_Mass_g_capped"]].describe()
Out[20]:
Body_Mass_g Body_Mass_g_capped
count 344.000000 344.000000
mean 4200.872093 4199.604651
std 799.696532 796.664777
min 2700.000000 2700.000000
25% 3550.000000 3550.000000
50% 4050.000000 4050.000000
75% 4750.000000 4750.000000
max 6300.000000 5978.500000
Feature Engineering
In [21]:
penguin_clean["Body Mass(kg)"] = penguin_clean["Body_Mass_g"]/1000
Type here to search
23°C
ENG
23:51
23-09-2025

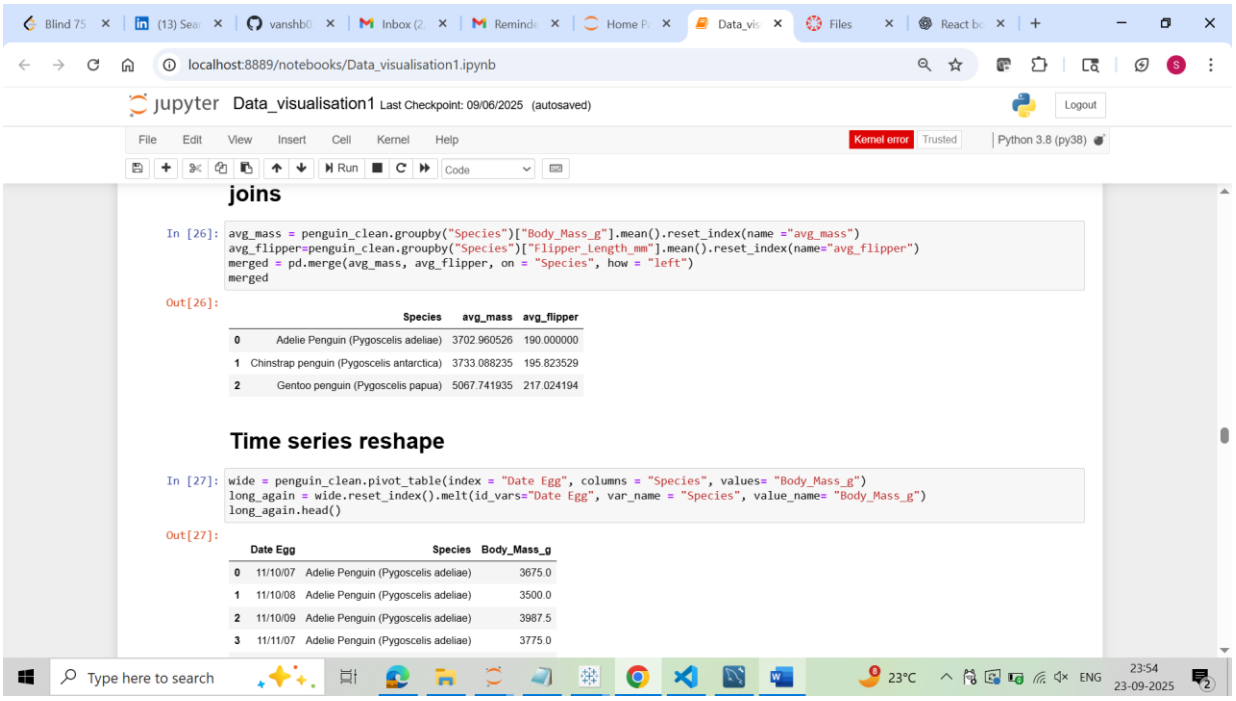

Blind 75
(13) Sea
vanshb0
Inbox (2
Remind
Home P
Data_vis
Files
React bc
localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error
Trusted
Python 3.8 (py38)
Run
Code
Feature Engineering
In [21]:
penguin_clean["Body Mass(kg)"] = penguin_clean["Body_Mass_g"]/1000
penguin_clean[["Body Mass(kg)"]]
Out[21]:
Body Mass(kg)
0 3.75
1 3.80
2 3.25
3 4.05
4 3.45
... ...
339 4.05
340 4.85
341 5.75
342 5.20
343 5.40
344 rows × 1 columns
In [22]:
penguin_clean["Flipper_per_Culmen"] = penguin_clean["Flipper_Length_mm"]/penguin_clean["Culmen_Depth_mm"]
penguin_clean[["Flipper_per_Culmen"]]
Type here to search
23°C
ENG
23:51
23-09-2025


Blind 75
(13) Sea
vanshb0
Inbox (2
Remind
Home P
Data_vis
Files
React bc
localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error
Trusted
Python 3.8 (py38)
Run
Code
344 rows × 1 columns
In [22]:
penguin_clean["Flipper_per_Culmen"] = penguin_clean["Flipper_Length_mm"]/penguin_clean["Culmen_Depth_mm"]
penguin_clean[["Flipper_per_Culmen"]]
Out[22]:
Flipper_per_Culmen
0 9.679144
1 10.689655
2 10.833333
3 11.387283
4 10.000000
... ...
339 11.387283
340 15.034965
341 14.140127
342 14.324324
343 13.229814
344 rows × 1 columns
Groupby and aggregate
Type here to search
23°C
ENG
23:53
23-09-2025


Blind 75
(13) Sea
vanshb0
Inbox (2
Remind
Home P
Data_vis
Files
React bc
localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error
Trusted
Python 3.8 (py38)
Run
Code
Groupby and aggregate
Comput average survival rates grouped by class and sex. GroupBy operations are central to data prep, letting us summarize and check trends quickly.
penguin_clean.groupby("Species") → creates a grouped object (by species).
["Body_Mass_g"] → tells pandas which numeric column you want to aggregate (take the mean of).
In [23]:
#Average body mass per species
Average_body_mass_per_species=(
penguin_clean
.groupby("Species")["Body_Mass_g"]
.mean()
.rename("body_mass")
.reset_index()
.sort_values(["Species"])
)
Average_body_mass_per_species
Out[23]:
Species body_mass
0 Adelie Penguin (Pygoscelis adeliae) 3702.960526
1 Chinstrap penguin (Pygoscelis antarctica) 3733.088235
2 Gentoo penguin (Pygoscelis papua) 5067.741935
In [24]:
avg_mass_species_sex = (
penguin_clean
.groupby(["Species", "Sex"])["Body_Mass_g"]
Type here to search
23°C
ENG
23:53
23-09-2025


Blind 75
(13) Sea
vanshb0
Inbox (2
Remind
Home P
Data_vis
Files
React bc
localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error
Trusted
Python 3.8 (py38)
Run
Code
In [24]:
avg_mass_species_sex = (
penguin_clean
.groupby(["Species", "Sex"])["Body_Mass_g"]
.mean()
.reset_index(name="mean_mass_g")
)
avg_mass_species_sex
Out[24]:
Species Sex mean_mass_g
0 Adelie Penguin (Pygoscelis adeliae) . NaN
1 Adelie Penguin (Pygoscelis adeliae) FEMALE 3378.040541
2 Adelie Penguin (Pygoscelis adeliae) MALE 4011.217949
3 Chinstrap penguin (Pygoscelis antarctica) . NaN
4 Chinstrap penguin (Pygoscelis antarctica) FEMALE 3527.205882
5 Chinstrap penguin (Pygoscelis antarctica) MALE 3938.970588
6 Gentoo penguin (Pygoscelis papua) . 4875.000000
7 Gentoo penguin (Pygoscelis papua) FEMALE 4669.067797
8 Gentoo penguin (Pygoscelis papua) MALE 5438.281250
#for my reference
#A pivot table is a way to summarize data in a grid by grouping rows and columns.
#Key difference
Type here to search
23°C
ENG
23:54
23-09-2025


Blind 75
(13) Sea
vanshb0
Inbox (2
Remind
Home P
Data_vis
Files
React bc
localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error
Trusted
Python 3.8 (py38)
Run
Code
Pivot tables
In [ ]:
In [25]:
pivot_table = avg_mass_species_sex.pivot(
index="Species",
columns="Sex",
values="mean_mass_g"
).round(3)
pivot_table
Out[25]:
Sex . FEMALE MALE
Species
Adelie Penguin (Pygoscelis adeliae) NaN 3378.041 4011.218
Chinstrap penguin (Pygoscelis antarctica) NaN 3527.206 3938.971
Gentoo penguin (Pygoscelis papua) 4875.0 4669.068 5438.281
joins
In [26]:
avg_mass = penguin_clean.groupby("Species")["Body_Mass_g"].mean().reset_index(name ="avg_mass")
Type here to search
23°C
ENG
23:54
23-09-2025


Blind 75
(13) Sea
vanshb0
Inbox (2
Remind
Home P
Data_vis
Files
React bc
localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error
Trusted
Python 3.8 (py38)
Run
Code
joins
In [26]:
avg_mass = penguin_clean.groupby("Species")["Body_Mass_g"].mean().reset_index(name ="avg_mass")
avg_flipper=penguin_clean.groupby("Species")["Flipper_Length_mm"].mean().reset_index(name="avg_flipper")
merged = pd.merge(avg_mass, avg_flipper, on = "Species", how = "left")
merged
Out[26]:
Species avg_mass avg_flipper
0 Adelie Penguin (Pygoscelis adeliae) 3702.960526 190.000000
1 Chinstrap penguin (Pygoscelis antarctica) 3733.088235 195.823529
2 Gentoo penguin (Pygoscelis papua) 5067.741935 217.024194
Time series reshape
In [27]:
wide = penguin_clean.pivot_table(index = "Date Egg", columns = "Species", values= "Body_Mass_g")
long_again = wide.reset_index().melt(id_vars="Date Egg", var_name = "Species", value_name= "Body_Mass_g")
long_again.head()
Out[27]:
Date Egg Species Body_Mass_g
0 11/10/07 Adelie Penguin (Pygoscelis adeliae) 3675.0
1 11/10/08 Adelie Penguin (Pygoscelis adeliae) 3500.0
2 11/10/09 Adelie Penguin (Pygoscelis adeliae) 3987.5
3 11/11/07 Adelie Penguin (Pygoscelis adeliae) 3775.0
Type here to search
23°C
ENG
23:54
23-09-2025

localhost:8889/notebooks/Data_visualisation1.ipynb

jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)

Logout

File Edit View Insert Cell Kernel Help

Kernel error Trusted Python 3.8 (py38)

Run Code

Validation checks

In [28]:
```
assert penguin_clean["Body_Mass_g"].ge(0).all()
assert penguin_clean["Flipper_Length_mm"].ge(0).all()
print("Basic checks passed ✅")
```

```
Basic checks passed ✅
```

Plotting

Line Chart ¶

In [29]:
```
import matplotlib.pyplot as plt
avg_mass = penguin_clean.groupby("Species")["Body_Mass_g"].mean().reset_index()

fig, ax = plt.subplots()
ax.plot(avg_mass["Species"], avg_mass["Body_Mass_g"], marker="o", linewidth=4)

ax.set_title("Average Penguin Body Mass by Species")
ax.set_xlabel("Species")
ax.set_ylabel("Body_Mass_g")

plt.tight_layout()  #automatically adjusts the spacing between subplots and around the edges so everything fits nicely.
plt.xticks(rotation=45) #rotates x labels to 45 as the names are too big
plt.show()
```

Type here to search 23°C ENG 23:54 23-09-2025

localhost:8889/notebooks/Data_visualisation1.ipynb

jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)

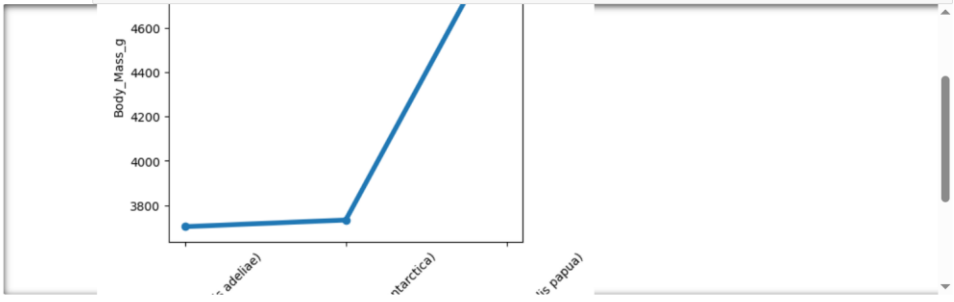
Logout

File Edit View Insert Cell Kernel Help

Kernel error Trusted Python 3.8 (py38)

Run Code

```
plt.xticks(rotation=45) #rotates x labels to 45 as the names are too big
plt.show()
```




```
The penguin dataset shows a clear distinction in body mass across species. Adelie and Chinstrap penguins have comparable average
body masses (~3.7 kg), while Gentoo penguins are significantly heavier (~5.1 kg). This makes body mass a reliable feature for
differentiating Gentoo penguins from the other two species.
```

Bar Graph

Type here to search 23°C ENG 23:55 23-09-2025

Blind 75 | (13) Sea | vanshb0 | Inbox (2 | Remind | Home P | Data_vis | Files | React bc

localhost:8889/notebooks/Data_visualisation1.ipynb

jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)

Logout

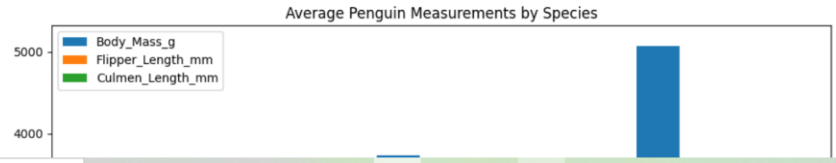
File Edit View Insert Cell Kernel Help

Kernel error Trusted Python 3.8 (py38)

Run Code

Bar Graph

In [30]:
```
import matplotlib.pyplot as plt

# Group and take the mean of each measurement per species
avg_traits = penguin_clean.groupby("Species")[["Body_Mass_g",
                                               "Flipper_Length_mm",
                                               "Culmen_Length_mm"]].mean()

# Now plot the DataFrame
ax = avg_traits.plot(kind="bar", figsize=(10,6))

ax.set_title("Average Penguin Measurements by Species")
ax.set_ylabel("Average Value")
plt.xticks(rotation=0)
plt.tight_layout()
plt.show()
```



Type here to search 23°C ENG 23:55 23-09-2025

Blind 75 | (13) Sea | vanshb0 | Inbox (2 | Remind | Home P | Data_vis | Files | React bc

localhost:8889/notebooks/Data_visualisation1.ipynb

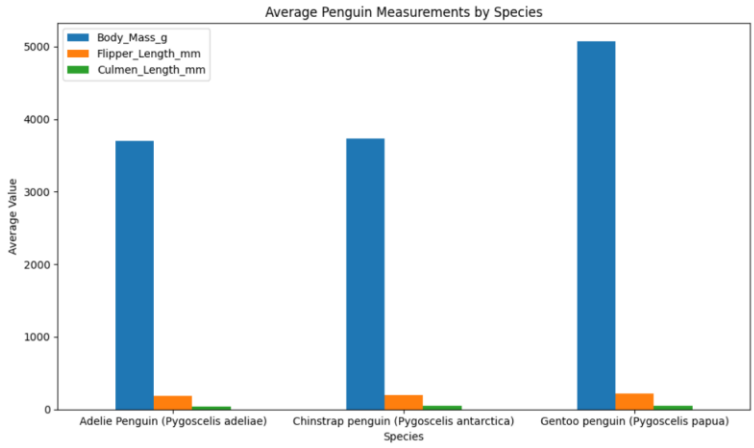
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)

Logout

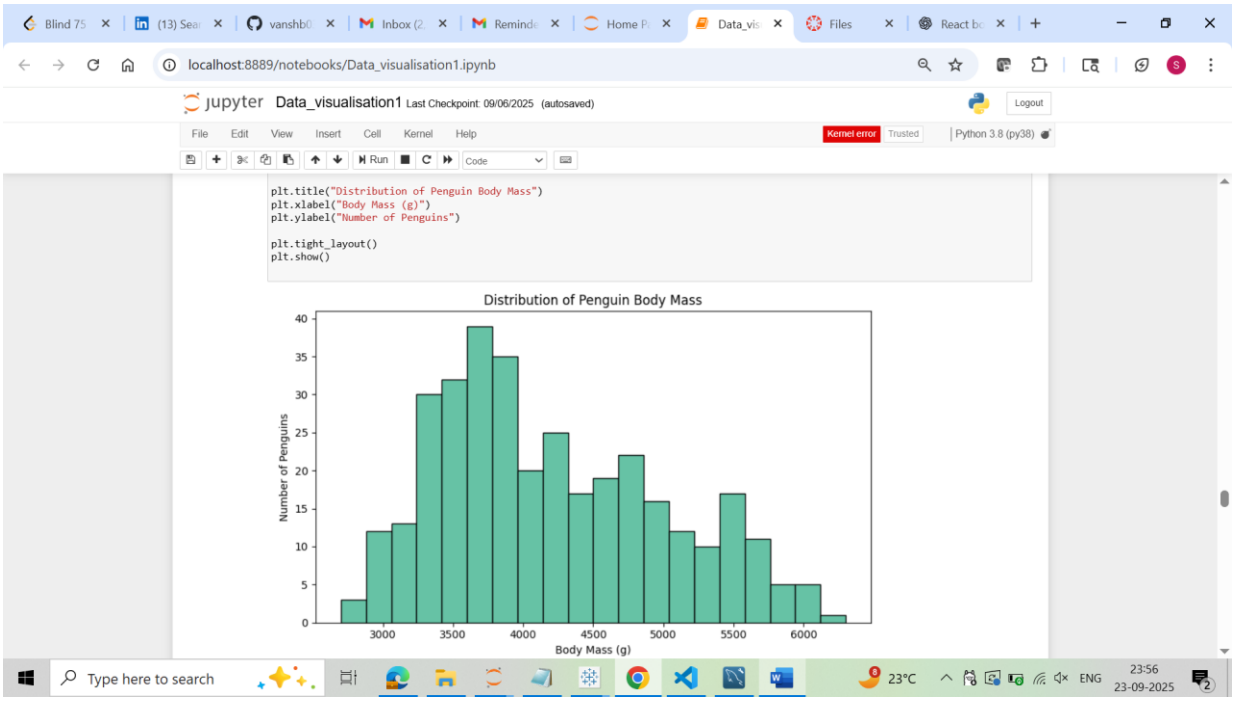
File Edit View Insert Cell Kernel Help

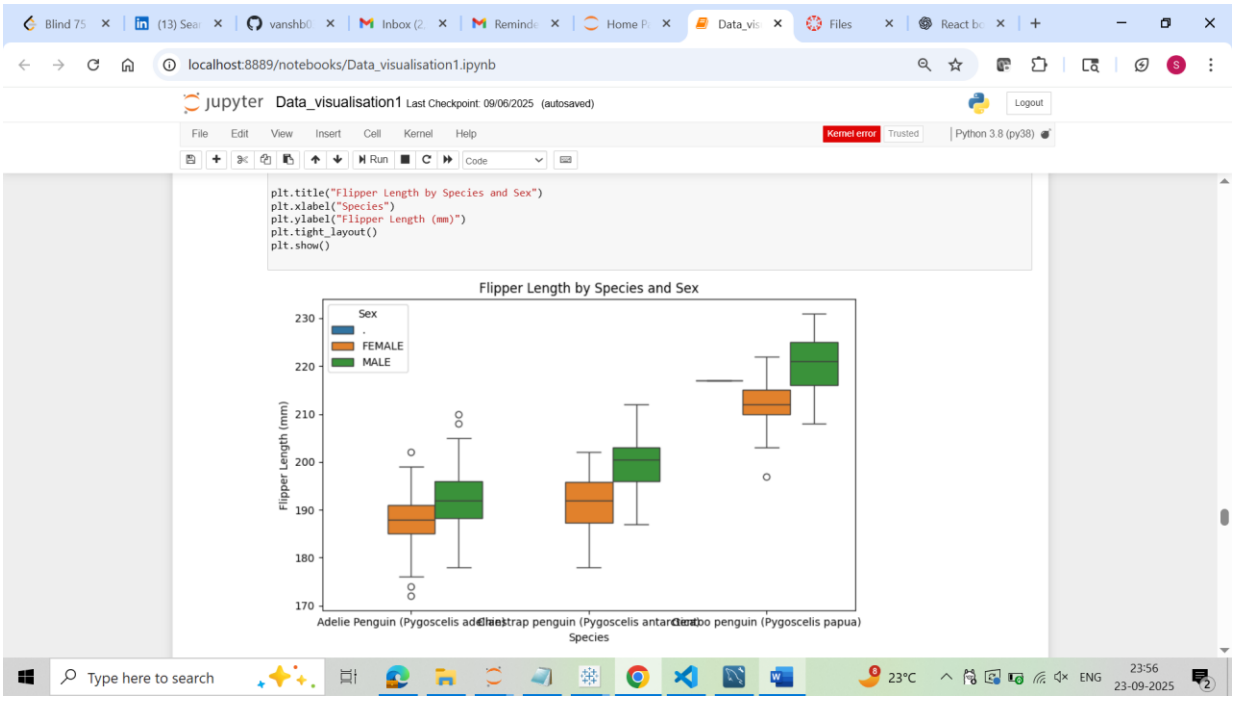
Kernel error Trusted Python 3.8 (py38)

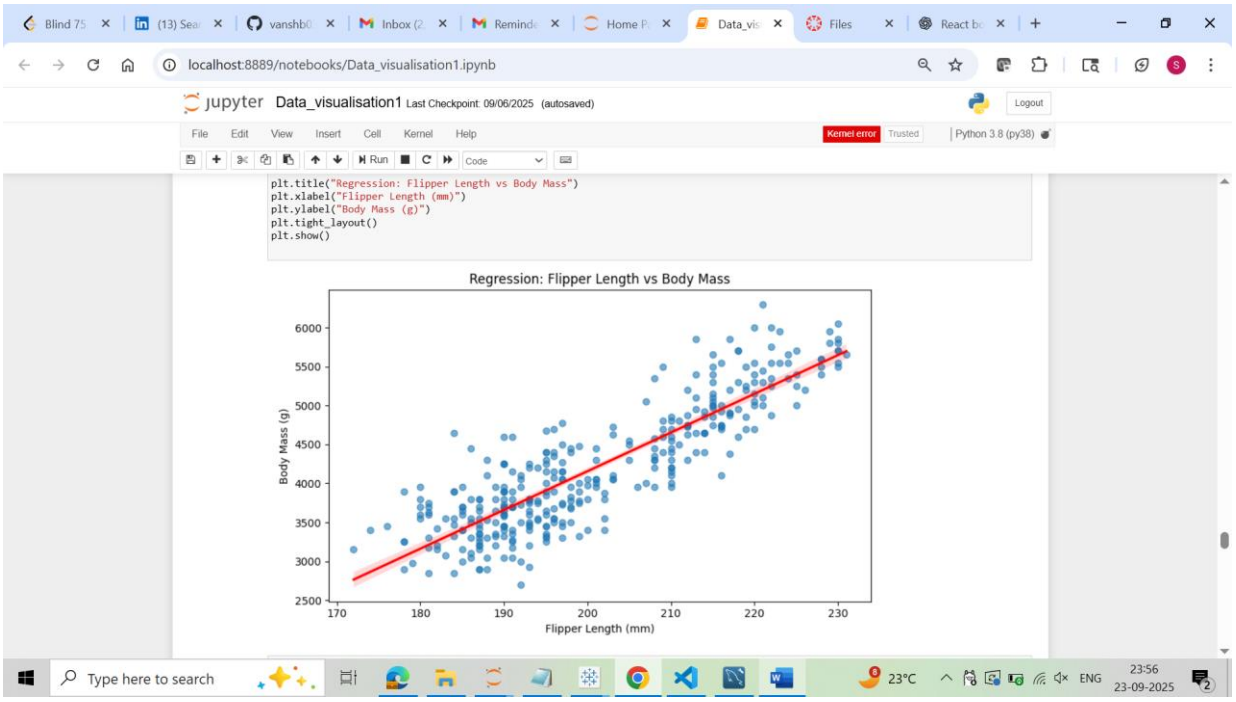
Run Code

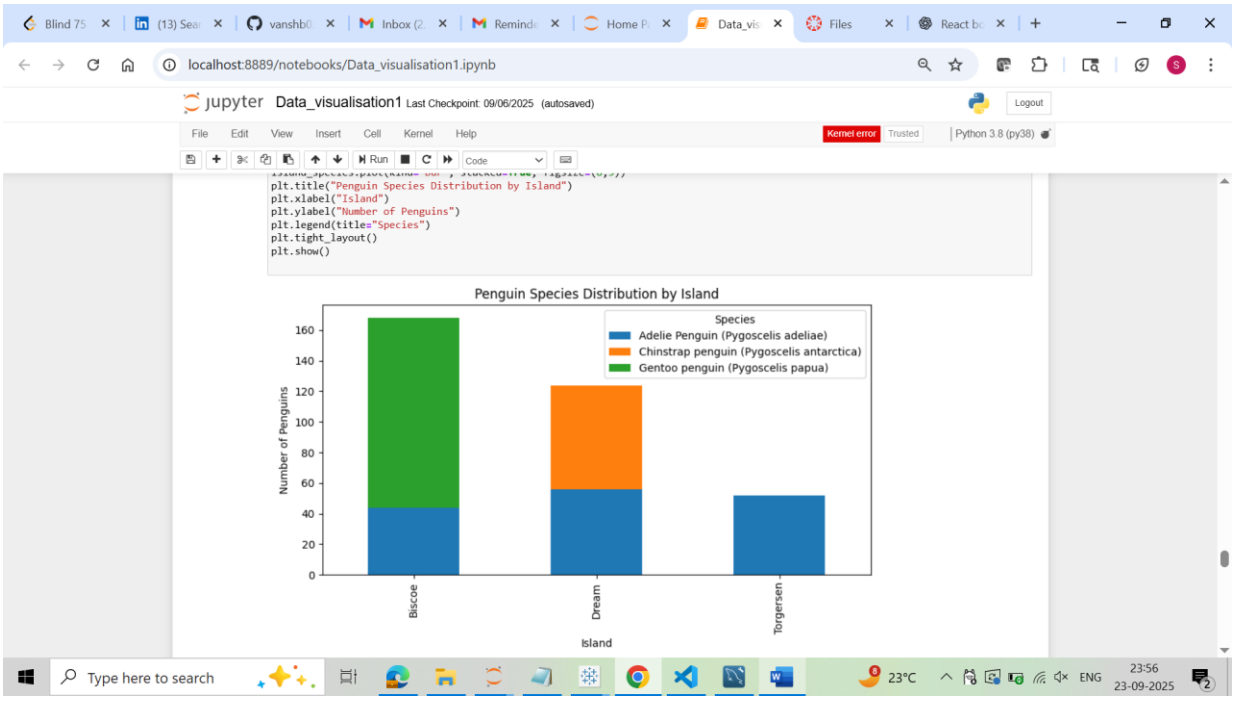


Type here to search 23°C ENG 23:55 23-09-2025


localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error Trusted Python 3.8 (py38)
Run Code
plt.title("Distribution of Penguin Body Mass")
plt.xlabel("Body Mass (g)")
plt.ylabel("Number of Penguins")
plt.tight_layout()
plt.show()
Distribution of Penguin Body Mass
Number of Penguins
Body Mass (g)
0
5
10
15
20
25
30
35
40
3000
3500
4000
4500
5000
5500
6000
Type here to search
23°C
ENG
23:56
23-09-2025


localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error Trusted Python 3.8 (py38)
Run Code
plt.title("Flipper Length by Species and Sex")
plt.xlabel("Species")
plt.ylabel("Flipper Length (mm)")
plt.tight_layout()
plt.show()
Flipper Length by Species and Sex
Sex
.
FEMALE
MALE
Flipper Length (mm)
170
180
190
200
210
220
230
Adelie Penguin (Pygoscelis adeliae)
Chinstrap penguin (Pygoscelis antarctica)
Gentoo penguin (Pygoscelis papua)
Species
Type here to search
23°C
ENG
23:56
23-09-2025


localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error Trusted Python 3.8 (py38)
Run Code
plt.title("Regression: Flipper Length vs Body Mass")
plt.xlabel("Flipper Length (mm)")
plt.ylabel("Body Mass (g)")
plt.tight_layout()
plt.show()
Regression: Flipper Length vs Body Mass
Body Mass (g)
Flipper Length (mm)
2500
3000
3500
4000
4500
5000
5500
6000
170
180
190
200
210
220
230
Type here to search
23°C
ENG
23:56
23-09-2025


localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Kernel error Trusted Python 3.8 (py38)
Run Code
plt.title("Penguin Species Distribution by Island")
plt.xlabel("Island")
plt.ylabel("Number of Penguins")
plt.legend(title="Species")
plt.tight_layout()
plt.show()
Penguin Species Distribution by Island
Species
Adelie Penguin (Pygoscelis adeliae)
Chinstrap penguin (Pygoscelis antarctica)
Gentoo penguin (Pygoscelis papua)
Number of Penguins
Island
Biscoe
Dream
Torgersen
0
20
40
60
80
100
120
140
160
Type here to search
23°C
ENG
23:56
23-09-2025

Blind 75 | (13) Sea | vanshb0 | Inbox (2 | Remind | Home P | Data_vis | Files | React bc

localhost:8889/notebooks/Data_visualisation1.ipynb

jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)

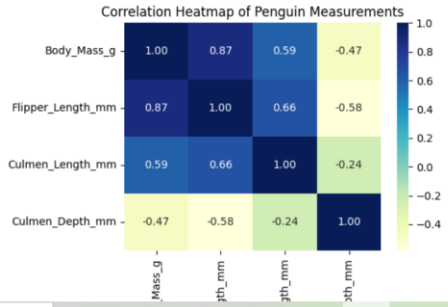
Logout

File Edit View Insert Cell Kernel Help

Kernel error Trusted Python 3.8 (py38)

Run Code

In [36]:
```
import seaborn as sns
import matplotlib.pyplot as plt

# Compute correlation matrix
num_cols = ["Body_Mass_g", "Flipper_Length_mm", "Culmen_Length_mm", "Culmen_Depth_mm"]
corr = penguin_clean[num_cols].corr()

plt.figure(figsize=(6,5))
sns.heatmap(corr, annot=True, cmap="YlGnBu", fmt=".2f")
plt.title("Correlation Heatmap of Penguin Measurements")
plt.tight_layout()
plt.show()
```



Type here to search

23°C ENG 23:57 23-09-2025

Blind 75 | (13) Sea | vanshb0 | Inbox (2 | Remind | Home P | Data_vis | Files | React bc

localhost:8889/notebooks/Data_visualisation1.ipynb

jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)

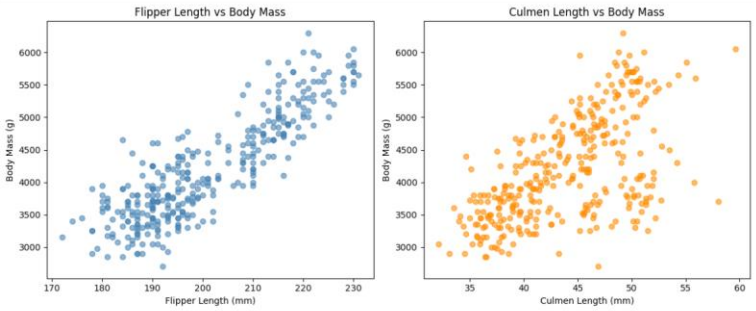
Logout

File Edit View Insert Cell Kernel Help

Kernel error Trusted Python 3.8 (py38)

Run Code

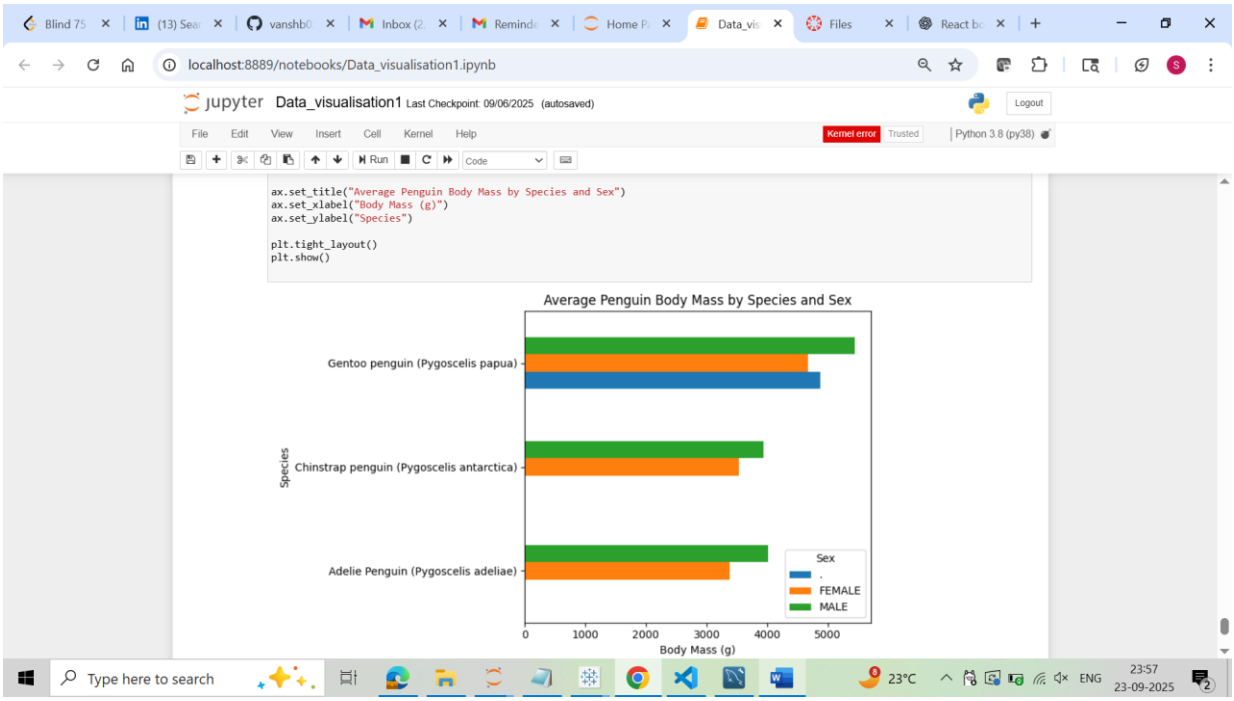
```
    alpha=0.6,
    c="darkorange"
)
ax2.set_title("Culmen Length vs Body Mass")
ax2.set_xlabel("Culmen Length (mm)")
ax2.set_ylabel("Body Mass (g)")

plt.tight_layout()
plt.show()
```



In [38]:
```
avg_mass_sex = (
```

Type here to search

23°C ENG 23:57 23-09-2025


localhost:8889/notebooks/Data_visualisation1.ipynb
jupyter Data_visualisation1 Last Checkpoint: 09/06/2025 (autosaved)
File Edit View Insert Cell Kernel Help
Kernel error Trusted Python 3.8 (py38)
ax.set_title("Average Penguin Body Mass by Species and Sex")
ax.set_xlabel("Body Mass (g)")
ax.set_ylabel("Species")
plt.tight_layout()
plt.show()
Average Penguin Body Mass by Species and Sex
Gentoo penguin (Pygoscelis papua)
Chinstrap penguin (Pygoscelis antarctica)
Adelie Penguin (Pygoscelis adeliae)
Species
Sex
FEMALE
MALE
0
1000
2000
3000
4000
5000
Body Mass (g)
Type here to search
23°C
ENG
23:57
23-09-2025

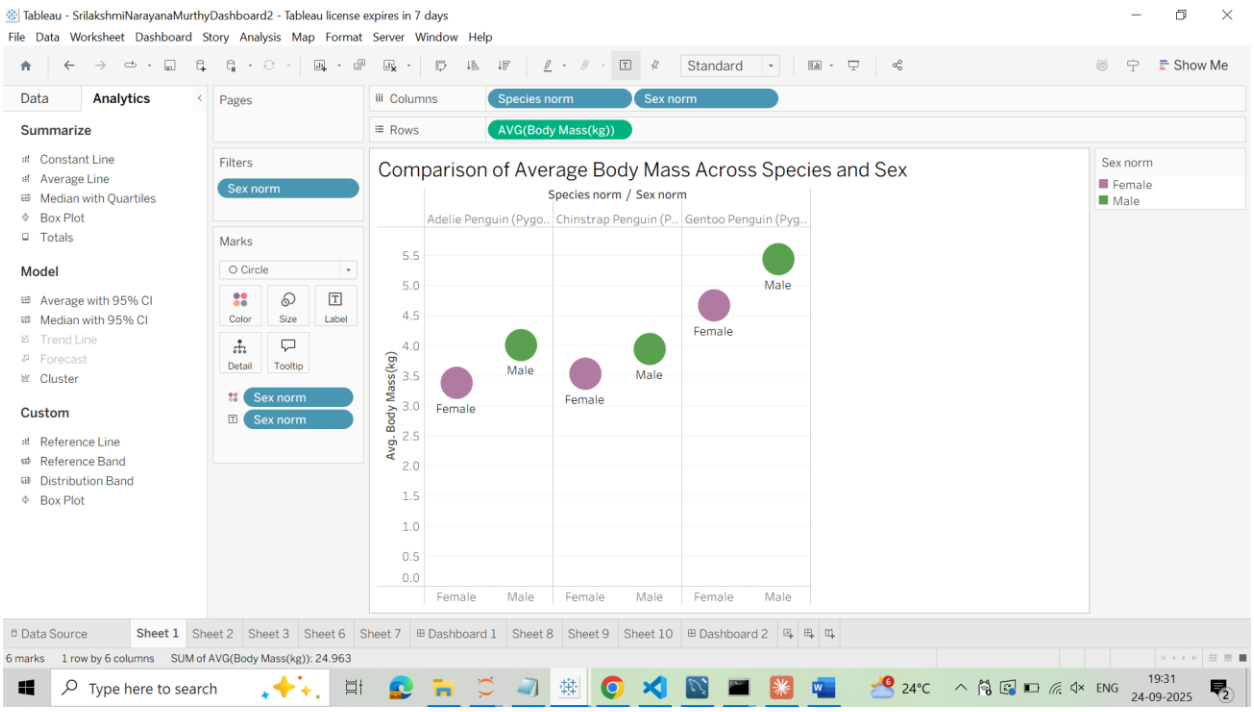


Tableau - SrilakshmiNarayanaMurthyDashboard2 - Tableau license expires in 7 days
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data
Analytics
Summarize
Constant Line
Average Line
Median with Quartiles
Box Plot
Totals
Model
Average with 95% CI
Median with 95% CI
Trend Line
Forecast
Cluster
Custom
Reference Line
Reference Band
Distribution Band
Box Plot
Pages
Filters
Sex norm
Marks
Circle
Color
Size
Label
Detail
Tooltip
Sex norm
Sex norm
Columns
Species norm
Sex norm
Rows
AVG(Body Mass(kg))
Comparison of Average Body Mass Across Species and Sex
Species norm / Sex norm
Adelie Penguin (Pygo..
Chinstrap Penguin (P..
Gentoo Penguin (Pyg..
Avg. Body Mass(kg)
Female
Male
Sex norm
Female
Male
Data Source
Sheet 1
Sheet 2
Sheet 3
Sheet 6
Sheet 7
Dashboard 1
Sheet 8
Sheet 9
Sheet 10
Dashboard 2
6 marks
1 row by 6 columns
SUM of AVG(Body Mass(kg)): 24.963
Type here to search
24°C
ENG
19:31
24-09-2025

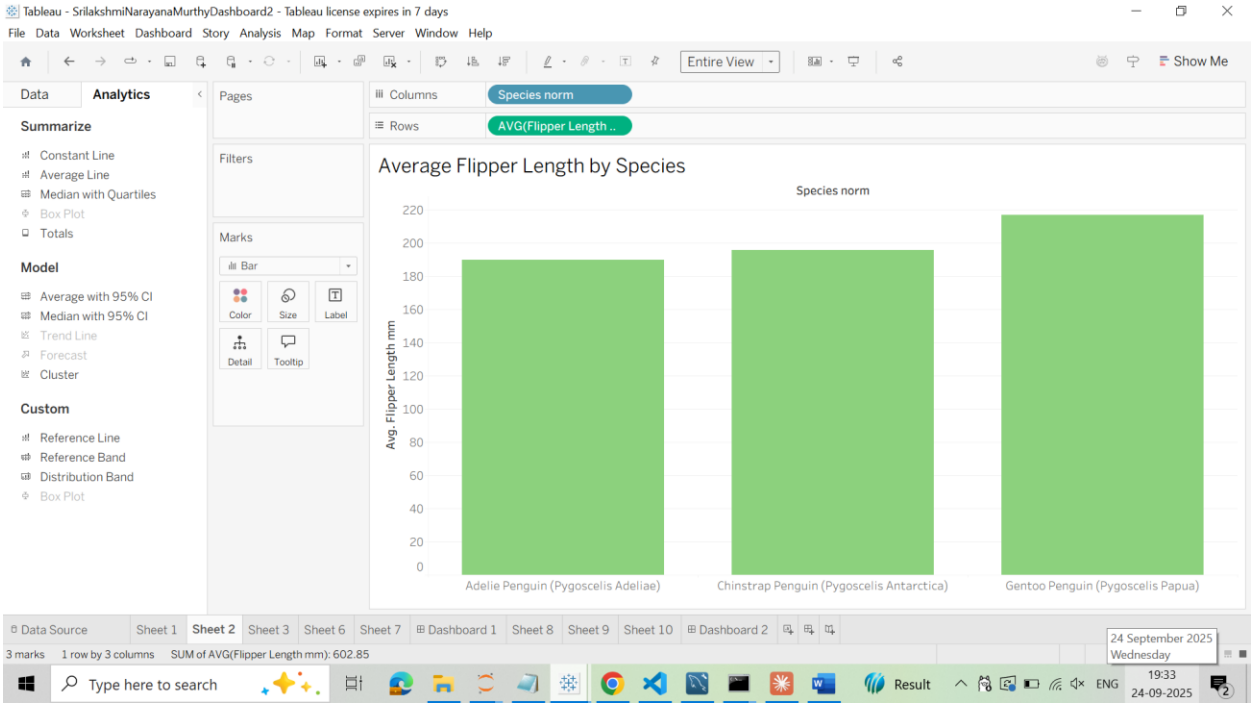


Tableau - SrilakshmiNarayanaMurthyDashboard2 - Tableau license expires in 7 days
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Entire View
Show Me
Data
Analytics
Summarize
Constant Line
Average Line
Median with Quartiles
Box Plot
Totals
Model
Average with 95% CI
Median with 95% CI
Trend Line
Forecast
Cluster
Custom
Reference Line
Reference Band
Distribution Band
Box Plot
Pages
Filters
Marks
Bar
Color
Size
Label
Detail
Tooltip
Columns
Species norm
Rows
AVG(Flipper Length ..
Average Flipper Length by Species
Species norm
Avg. Flipper Length mm
220
200
180
160
140
120
100
80
60
40
20
0
Adelie Penguin (Pygoscelis Adeliae)
Chinstrap Penguin (Pygoscelis Antarctica)
Gentoo Penguin (Pygoscelis Papua)
Data Source
Sheet 1
Sheet 2
Sheet 3
Sheet 6
Sheet 7
Dashboard 1
Sheet 8
Sheet 9
Sheet 10
Dashboard 2
3 marks
1 row by 3 columns
SUM of AVG(Flipper Length mm): 602.85
24 September 2025
Wednesday
Type here to search
Result
ENG
19:33
24-09-2025

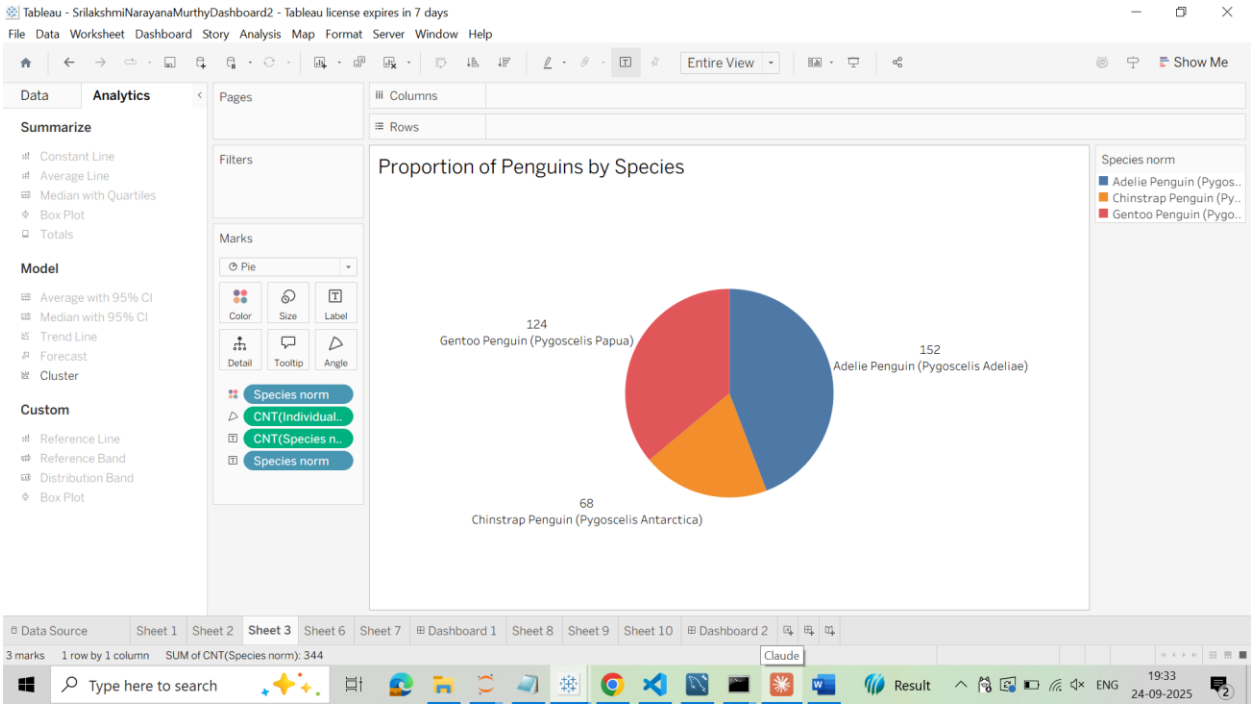


Tableau - SrilakshmiNarayanaMurthyDashboard2 - Tableau license expires in 7 days
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Entire View
Show Me
Data
Analytics
Summarize
Constant Line
Average Line
Median with Quartiles
Box Plot
Totals
Model
Average with 95% CI
Median with 95% CI
Trend Line
Forecast
Cluster
Custom
Reference Line
Reference Band
Distribution Band
Box Plot
Pages
Filters
Marks
Pie
Color
Size
Label
Detail
Tooltip
Angle
Species norm
CNT(Individual..
CNT(Species n..
Species norm
Columns
Rows
Proportion of Penguins by Species
Species norm
Adelie Penguin (Pygos..
Chinstrap Penguin (Py..
Gentoo Penguin (Pygo..
124
Gentoo Penguin (Pygoscelis Papua)
152
Adelie Penguin (Pygoscelis Adeliae)
68
Chinstrap Penguin (Pygoscelis Antarctica)
Data Source
Sheet 1
Sheet 2
Sheet 3
Sheet 6
Sheet 7
Dashboard 1
Sheet 8
Sheet 9
Sheet 10
Dashboard 2
3 marks
1 row by 1 column
SUM of CNT(Species norm): 344
Claude
Type here to search
Result
ENG
19:33
24-09-2025

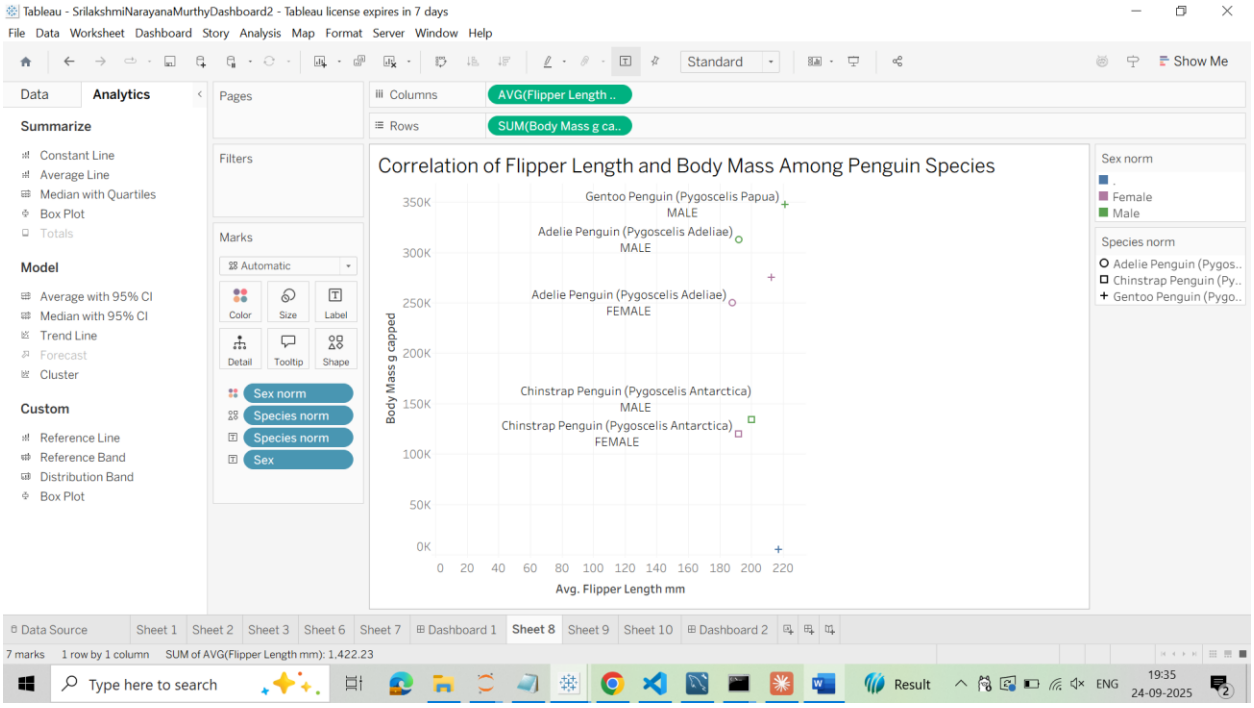


Tableau - SrilakshmiNarayanaMurthyDashboard2 - Tableau license expires in 7 days
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data
Analytics
Summarize
Constant Line
Average Line
Median with Quartiles
Box Plot
Totals
Model
Average with 95% CI
Median with 95% CI
Trend Line
Forecast
Cluster
Custom
Reference Line
Reference Band
Distribution Band
Box Plot
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Shape
Sex norm
Species norm
Species norm
Sex
Columns
AVG(Flipper Length ..
Rows
SUM(Body Mass g ca..
Correlation of Flipper Length and Body Mass Among Penguin Species
Gentoo Penguin (Pygoscelis Papua)
MALE
Adelie Penguin (Pygoscelis Adeliae)
MALE
Adelie Penguin (Pygoscelis Adeliae)
FEMALE
Chinstrap Penguin (Pygoscelis Antarctica)
MALE
Chinstrap Penguin (Pygoscelis Antarctica)
FEMALE
Body Mass g capped
350K
300K
250K
200K
150K
100K
50K
0K
0 20 40 60 80 100 120 140 160 180 200 220
Avg. Flipper Length mm
Sex norm
Female
Male
Species norm
Adelie Penguin (Pygos..
Chinstrap Penguin (Py..
Gentoo Penguin (Pygo..
Data Source
Sheet 1
Sheet 2
Sheet 3
Sheet 6
Sheet 7
Dashboard 1
Sheet 8
Sheet 9
Sheet 10
Dashboard 2
7 marks
1 row by 1 column
SUM of AVG(Flipper Length mm): 1,422.23
Type here to search
Result
ENG
19:35
24-09-2025

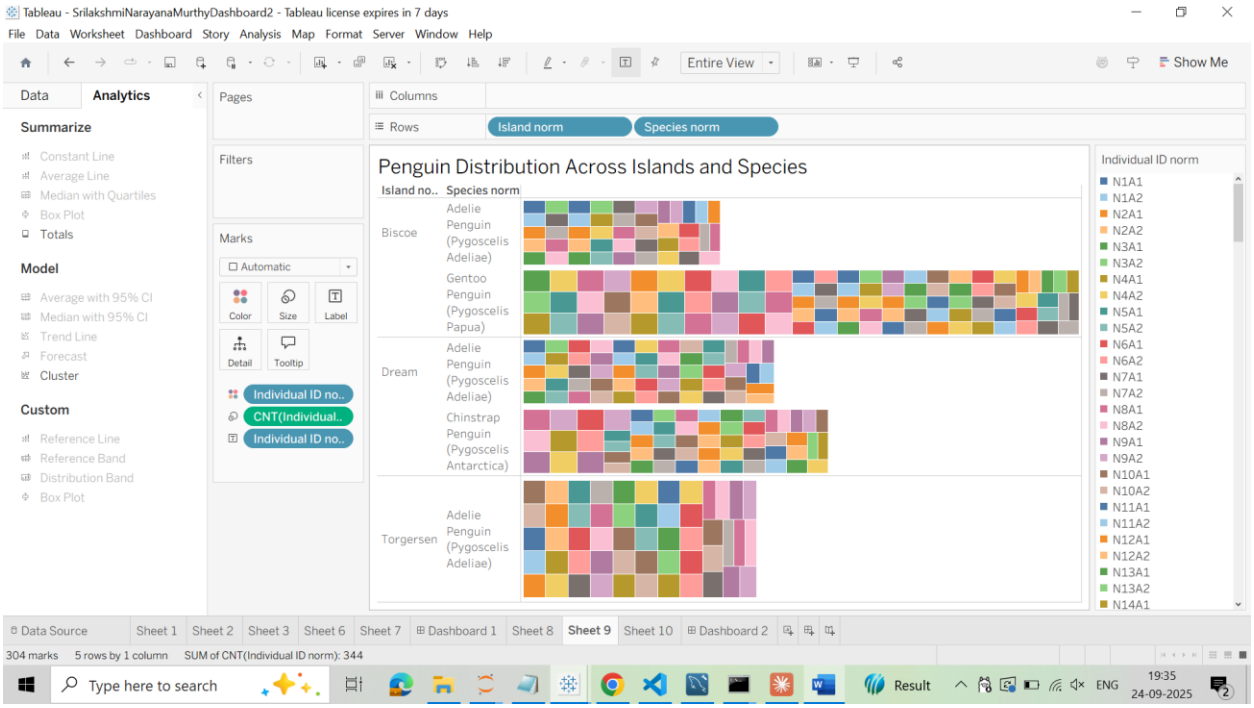


Tableau - SrilakshmiNarayanaMurthyDashboard2 - Tableau license expires in 7 days
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Entire View
Show Me
Data
Analytics
Summarize
Constant Line
Average Line
Median with Quartiles
Box Plot
Totals
Model
Average with 95% CI
Median with 95% CI
Trend Line
Forecast
Cluster
Custom
Reference Line
Reference Band
Distribution Band
Box Plot
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Individual ID no..
CNT(Individual..
Individual ID no..
Columns
Rows
Island norm
Species norm
Penguin Distribution Across Islands and Species
Island no..
Species norm
Biscoe
Adelie Penguin (Pygoscelis Adeliae)
Gentoo Penguin (Pygoscelis Papua)
Dream
Adelie Penguin (Pygoscelis Adeliae)
Chinstrap Penguin (Pygoscelis Antarctica)
Torgersen
Adelie Penguin (Pygoscelis Adeliae)
Individual ID norm
N1A1
N1A2
N2A1
N2A2
N3A1
N3A2
N4A1
N4A2
N5A1
N5A2
N6A1
N6A2
N7A1
N7A2
N8A1
N8A2
N9A1
N9A2
N10A1
N10A2
N11A1
N11A2
N12A1
N12A2
N13A1
N13A2
N14A1
Data Source
Sheet 1
Sheet 2
Sheet 3
Sheet 6
Sheet 7
Dashboard 1
Sheet 8
Sheet 9
Sheet 10
Dashboard 2
304 marks
5 rows by 1 column
SUM of CNT(Individual ID norm): 344
Type here to search
Result
ENG
19:35
24-09-2025

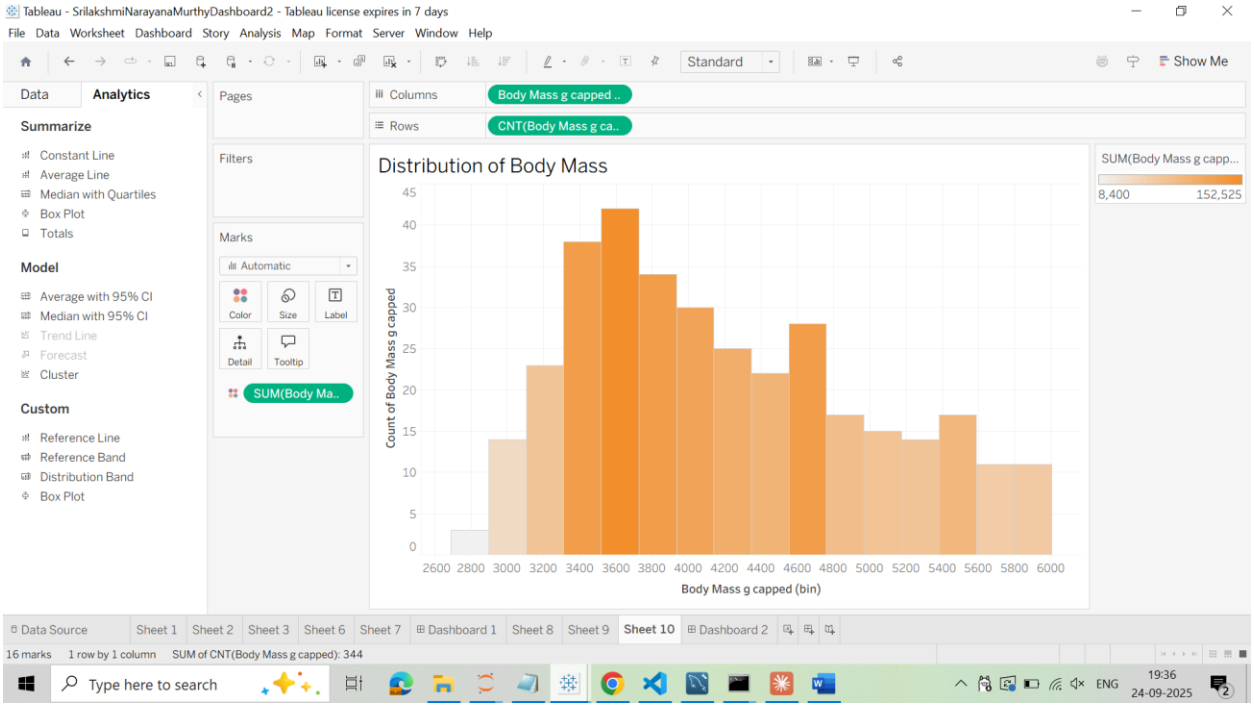


Tableau - SrilakshmiNarayanaMurthyDashboard2 - Tableau license expires in 7 days
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data
Analytics
Summarize
Constant Line
Average Line
Median with Quartiles
Box Plot
Totals
Model
Average with 95% CI
Median with 95% CI
Trend Line
Forecast
Cluster
Custom
Reference Line
Reference Band
Distribution Band
Box Plot
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
SUM(Body Ma..
Columns
Body Mass g capped ..
Rows
CNT(Body Mass g ca..
Distribution of Body Mass
Count of Body Mass g capped
Body Mass g capped (bin)
SUM(Body Mass g capp...
8,400
152,525
Data Source
Sheet 1
Sheet 2
Sheet 3
Sheet 6
Sheet 7
Dashboard 1
Sheet 8
Sheet 9
Sheet 10
Dashboard 2
16 marks
1 row by 1 column
SUM of CNT(Body Mass g capped): 344
Type here to search
ENG
19:36
24-09-2025

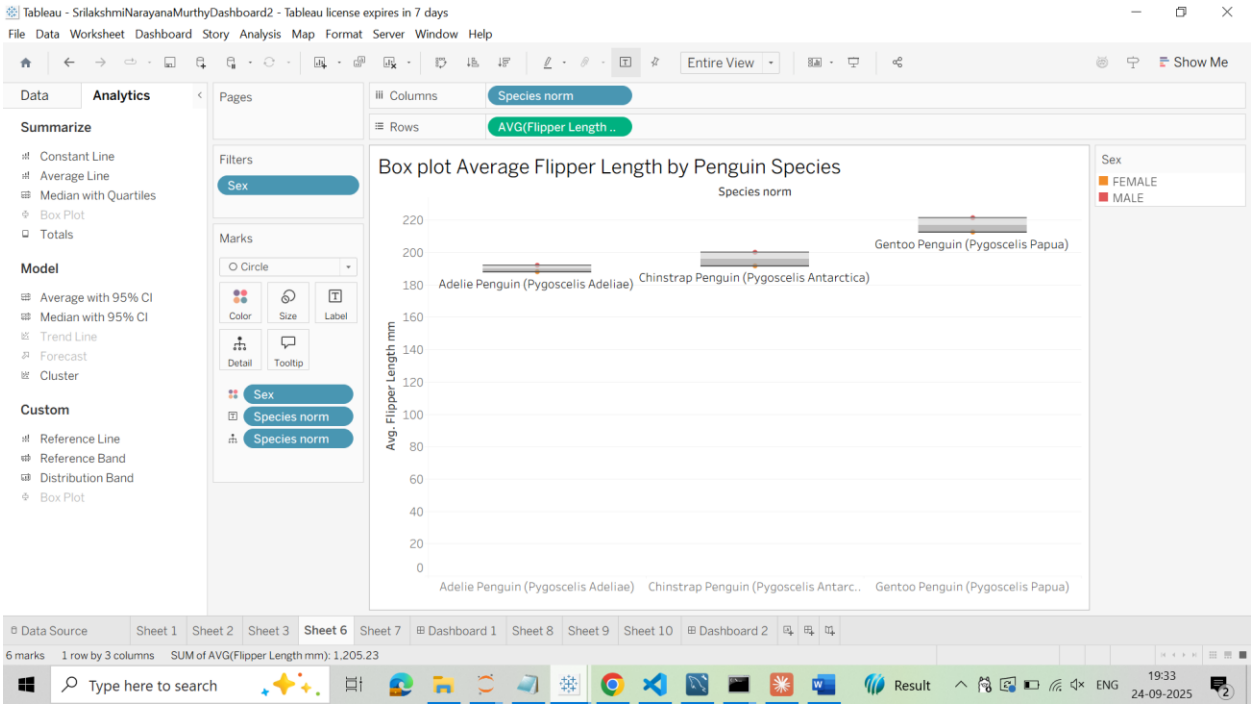


Tableau - SrilakshmiNarayanaMurthyDashboard2 - Tableau license expires in 7 days
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Entire View
Show Me
Data
Analytics
Summarize
Constant Line
Average Line
Median with Quartiles
Box Plot
Totals
Model
Average with 95% CI
Median with 95% CI
Trend Line
Forecast
Cluster
Custom
Reference Line
Reference Band
Distribution Band
Box Plot
Pages
Filters
Sex
Marks
Circle
Color
Size
Label
Detail
Tooltip
Sex
Species norm
Species norm
Columns
Species norm
Rows
AVG(Flipper Length ..
Box plot Average Flipper Length by Penguin Species
Species norm
Avg. Flipper Length mm
Adelie Penguin (Pygoscelis Adeliae)
Chinstrap Penguin (Pygoscelis Antarctica)
Gentoo Penguin (Pygoscelis Papua)
Sex
FEMALE
MALE
Data Source
Sheet 1
Sheet 2
Sheet 3
Sheet 6
Sheet 7
Dashboard 1
Sheet 8
Sheet 9
Sheet 10
Dashboard 2
6 marks
1 row by 3 columns
SUM of AVG(Flipper Length mm): 1,205.23
Type here to search
Result
ENG
19:33
24-09-2025

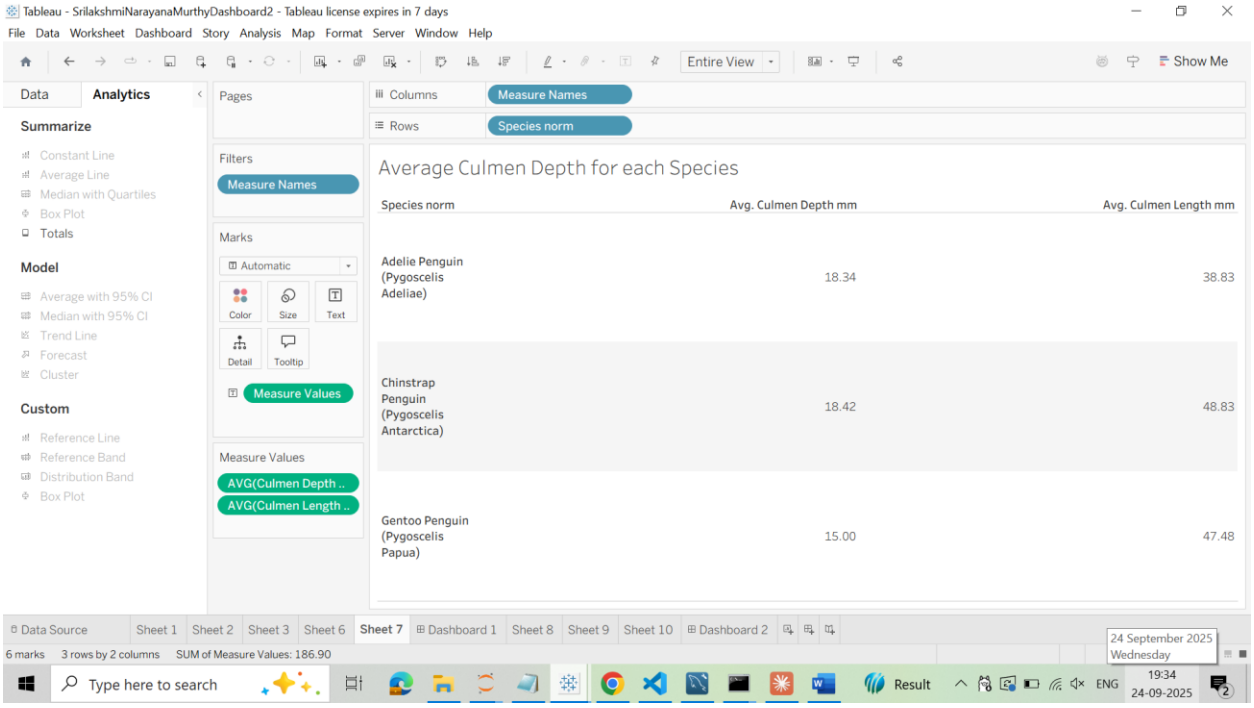


Tableau - SrilakshmiNarayanaMurthyDashboard2 - Tableau license expires in 7 days
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data Analytics
Summarize
Constant Line
Average Line
Median with Quartiles
Box Plot
Totals
Model
Average with 95% CI
Median with 95% CI
Trend Line
Forecast
Cluster
Custom
Reference Line
Reference Band
Distribution Band
Box Plot
Pages
Filters
Measure Names
Marks
Automatic
Color
Size
Text
Detail
Tooltip
Measure Values
Measure Values
AVG(Culmen Depth ..
AVG(Culmen Length ..
Columns
Measure Names
Rows
Species norm
Average Culmen Depth for each Species
Species norm
Avg. Culmen Depth mm
Avg. Culmen Length mm
Adelie Penguin (Pygoscelis Adeliae)
18.34
38.83
Chinstrap Penguin (Pygoscelis Antarctica)
18.42
48.83
Gentoo Penguin (Pygoscelis Papua)
15.00
47.48
Data Source
Sheet 1
Sheet 2
Sheet 3
Sheet 6
Sheet 7
Dashboard 1
Sheet 8
Sheet 9
Sheet 10
Dashboard 2
6 marks
3 rows by 2 columns
SUM of Measure Values: 186.90
24 September 2025
Wednesday
Type here to search
Result
ENG
19:34
24-09-2025

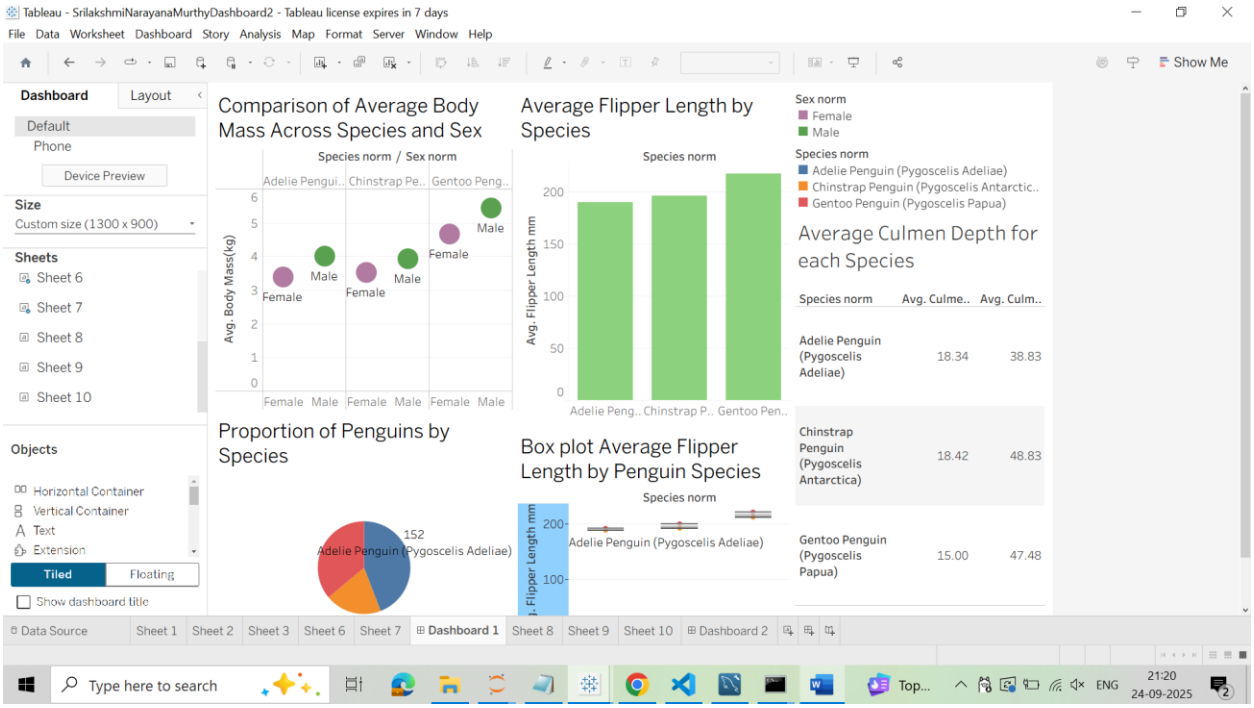


Tableau - SrilakshmiNarayanaMurthyDashboard2 - Tableau license expires in 7 days
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Dashboard
Layout
Default
Phone
Device Preview
Size
Custom size (1300 x 900)
Sheets
Sheet 6
Sheet 7
Sheet 8
Sheet 9
Sheet 10
Objects
Horizontal Container
Vertical Container
Text
Extension
Tiled
Floating
Show dashboard title
Comparison of Average Body Mass Across Species and Sex
Species norm / Sex norm
Adelie Pengui..
Chinstrap Pe..
Gentoo Peng..
Avg. Body Mass(kg)
Female
Male
Average Flipper Length by Species
Species norm
Avg. Flipper Length mm
Adelie Peng..
Chinstrap P..
Gentoo Pen..
Sex norm
Female
Male
Species norm
Adelie Penguin (Pygoscelis Adeliae)
Chinstrap Penguin (Pygoscelis Antarctic..
Gentoo Penguin (Pygoscelis Papua)
Average Culmen Depth for each Species
Species norm
Avg. Culme..
Avg. Culm..
Adelie Penguin (Pygoscelis Adeliae)
18.34
38.83
Chinstrap Penguin (Pygoscelis Antarctica)
18.42
48.83
Gentoo Penguin (Pygoscelis Papua)
15.00
47.48
Proportion of Penguins by Species
152
Adelie Penguin (Pygoscelis Adeliae)
Box plot Average Flipper Length by Penguin Species
Species norm
Adelie Penguin (Pygoscelis Adeliae)
Data Source
Sheet 1
Sheet 2
Sheet 3
Sheet 6
Sheet 7
Dashboard 1
Sheet 8
Sheet 9
Sheet 10
Dashboard 2
Type here to search
ENG
21:20
24-09-2025

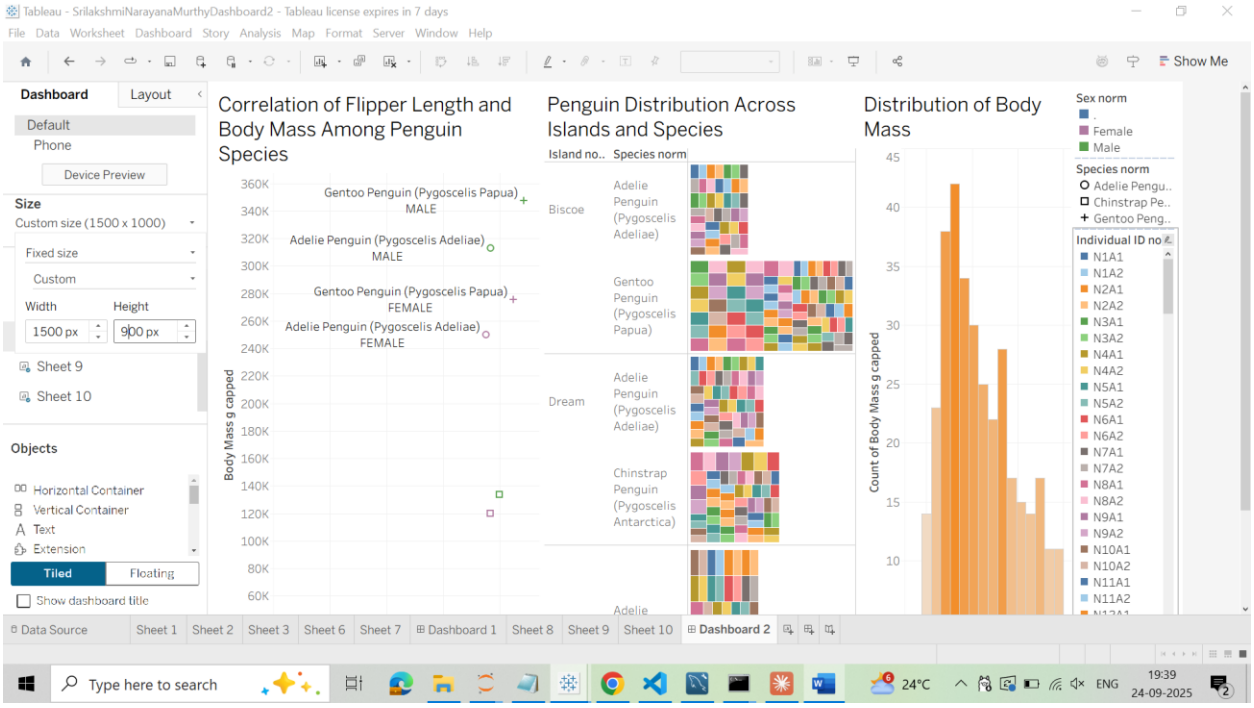


Tableau - SrilakshmiNarayanaMurthyDashboard2 - Tableau license expires in 7 days
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Dashboard
Layout
Default
Phone
Device Preview
Size
Custom size (1500 x 1000)
Fixed size
Custom
Width
Height
1500 px
900 px
Sheet 9
Sheet 10
Objects
Horizontal Container
Vertical Container
Text
Extension
Tiled
Floating
Show dashboard title
Correlation of Flipper Length and Body Mass Among Penguin Species
Gentoo Penguin (Pygoscelis Papua) MALE
Adelie Penguin (Pygoscelis Adeliae) MALE
Gentoo Penguin (Pygoscelis Papua) FEMALE
Adelie Penguin (Pygoscelis Adeliae) FEMALE
Body Mass g capped
Penguin Distribution Across Islands and Species
Island no..
Species norm
Biscoe
Adelie Penguin (Pygoscelis Adeliae)
Gentoo Penguin (Pygoscelis Papua)
Dream
Adelie Penguin (Pygoscelis Adeliae)
Chinstrap Penguin (Pygoscelis Antarctica)
Distribution of Body Mass
Count of Body Mass g capped
Sex norm
Female
Male
Species norm
Adelie Pengu..
Chinstrap Pe..
Gentoo Peng..
Individual ID no
Data Source
Sheet 1
Sheet 2
Sheet 3
Sheet 6
Sheet 7
Dashboard 1
Sheet 8
Sheet 9
Sheet 10
Dashboard 2
Type here to search
24°C
ENG
19:39
24-09-2025