Here's the analysis of the code with potential viva questions and answers:

---

### \*\*Code Explanation\*\*

1. \*\*Loading Data\*\*

```python

iris = sns.load\_dataset('iris')

```

- Loads the classic Iris dataset (150 flowers, 4 features, 3 species)

- Columns: sepal\_length, sepal\_width, petal\_length, petal\_width, species

2. \*\*Initial Data Inspection\*\*

```python

print(iris.head()) # First 5 rows

print(iris.dtypes) # Data types (float64 for measurements, category for species)

```

3. \*\*Histogram Grid\*\*

```python

iris.hist(figsize=(10,8), edgecolor='black')

```

- Creates histograms for all numeric columns

- Shows distribution of each feature

- \*\*Key Insight\*\*: Petal measurements are bimodal (two peaks), sepals are more normally distributed

4. \*\*Boxplot Grid\*\*

```python

sns.boxplot(y=iris[feature])

```

- Visualizes spread and outliers

- \*\*Key Insight\*\*: Petal lengths have clear outliers, sepal width shows most variability

5. \*\*Custom Histogram Grid\*\*

```python

plt.hist(iris[feature], bins=15)

```

- Manual recreation of histograms with consistent styling

- \*\*Key Insight\*\*: Petal measurements separate species clusters better than sepal measurements

---

### \*\*Viva Questions & Answers\*\*

\*\*Q1: Why use both histograms and boxplots?\*\*

A1:

- Histograms show distribution shape (normal/skewed/bimodal)

- Boxplots highlight spread, quartiles, and outliers

- Together they give complementary information

\*\*Q2: What does `figsize=(10,8)` do?\*\*

A2: Sets figure dimensions to 10" wide × 8" tall for better readability

\*\*Q3: Why use `plt.tight\_layout()`?\*\*

A3: Automatically adjusts subplot spacing to prevent label overlapping

\*\*Q4: How would you handle the species column?\*\*

A4: Could add:

```python

sns.pairplot(iris, hue='species') # Color by species

```

\*\*Q5: What patterns do you see in petal measurements?\*\*

A5:

- Clear separation between species

- Setosa (small petals), Versicolor (medium), Virginica (large)

- Bimodal distribution in petal length/width

\*\*Q6: What does edgecolor='black' do in histograms?\*\*

A6: Adds black borders to histogram bars for better visual separation

\*\*Q7: How to interpret sepal width boxplot?\*\*

A7:

- Median line around 3.0

- IQR (middle 50%) between ~2.8-3.3

- Several outliers below 2.2 and above 4.0

\*\*Q8: Why 15 bins in histograms?\*\*

A8: Trial value - enough to show detail without being too granular. Rule of thumb: √n ≈ 12 for 150 samples

\*\*Q9: What machine learning task suits this data?\*\*

A9: Classification (predicting species from measurements)

\*\*Q10: How would you improve this analysis?\*\*

A10: Add:

- Correlation matrix (`iris.corr()`)

- Species-wise comparisons

- Scatterplots showing feature relationships

---

### \*\*Key Insights\*\*

1. \*\*Feature Distributions\*\*:

- Sepal measurements: More normal distribution

- Petal measurements: Bimodal (reflects species differences)

2. \*\*Outliers\*\*:

- Sepal width has most outliers

- Petal lengths are clean except few cases

3. \*\*Species Separation\*\*:

- Petal features better for classification

- Sepal features overlap more between species

---

### \*\*Syllabus Connections\*\*

1. \*\*Exploratory Data Analysis (EDA)\*\*: Core data science skill shown here

2. \*\*Data Visualization\*\*: Matplotlib/Seaborn proficiency

3. \*\*Descriptive Statistics\*\*: Understanding distributions

4. \*\*Feature Engineering\*\*: Identifying important features

Would you like me to demonstrate any specific enhancement to this analysis? 😊