Here is your **Viva Explanation Markdown File** (you can copy & save as
`kmeans_sales.md`):

K-Means Clustering on Sales Data (Viva Explanation)

1. Objective
The aim of this practical is to group sales records into clusters using the **K-Means Cluste
Clustering helps us find hidden patterns or groups in data.

---

2. Libraries Used
Library	Purpose
pandas	For loading and managing data
matplotlib	For plotting graphs
sklearn.cluster (KMeans)	To perform clustering
sklearn.preprocessing (StandardScaler)	To normalize data

---

3. Dataset Used
We load `sales_data_sample.csv` and only use two columns:

- `ORDERLINENUMBER` (represents order/entry number)
- `SALES` (represents sales amount)

```python
df = df[['ORDERLINENUMBER', 'SALES']]
```

---

4. Data Normalization

Since the values in `SALES` are much larger, we scale the data using **Standard-
Scaler** to avoid bias.

```
scaled_values = scaler.fit_transform(df.values)
```

---

5. Finding Optimal Number of Clusters (Elbow Method)

We run K-Means for **1 to 10 clusters** and store the **WCSS (Within Cluster
Sum of Squares)**.

```python
for i in range(1, 11):
    model = KMeans(n_clusters=i, init='k-means++')
    model.fit_predict(scaled_values)
    wcss.append(model.inertia_)
```

Then we plot the **Elbow Curve**:

```python
plt.plot(range(1, 11), wcss, 'ro-')
```

Interpretation:

The point where the curve bends ("elbow") indicates the best number of clusters.

---

6. Applying K-Means

We choose **7 clusters** (based on the elbow point).

```python
model = KMeans(n_clusters=7, init='k-means++')
clusters = model.fit_predict(scaled_values)
df['cluster'] = clusters
```

---

7. Visualization

We visualize clusters on a scatter plot:

```python
plt.scatter(df['ORDERLINENUMBER'], df['SALES'], c=df['cluster'])
plt.show()
```

Each color in the graph represents one cluster.

---

8. Result

- Data points are divided into **7 clusters**.
- Similar sales patterns are grouped together.
- Helps in identifying customer purchasing patterns or sales categories.

---

9. Conclusion

K-Means clustering helps group similar data without any labeled output. By using the Elbow Method, we identify the optimal number of clusters and classify sales data into meaningful groups.

If you want, I can now also **generate a downloadable `.md` file** automatically. Just reply: **"Yes create file"**