```
# ================================
# LOAD ALL DATASETS
# ================================

import pandas as pd
from sklearn.preprocessing import StandardScaler
from scipy.cluster.hierarchy import linkage, fcluster

# Upload files in Colab:
from google.colab import files
try:
    print("Please ensure the file upload dialog appears in your brow
    uploaded = files.upload()
    print("Files uploaded successfully (or skipped if already presen
except Exception as e:
    print(f"An error occurred during file upload: {type(e).__name__}
    print("This often indicates a browser-related issue with the int
    print("Please try the following:")
    print("1. Refresh your browser tab.")
    print("2. Restart the Colab runtime (Runtime -> Restart runtime.
    print("3. Ensure the file upload dialog appears and you select t
    print("If the files are already in your Colab environment, you c

# After uploading, run:
sku = pd.read_csv('DATA_2.01_SKU.csv')
hr  = pd.read_csv('DATA_2.02_HR.csv')
tel = pd.read_csv('DATA_2.03_Telco.csv')

# ================================
# QUESTION 4 — HR CLUSTERING (2 clusters)
# ================================

X = hr[['S','LPE','NP']].copy()   # Satisfaction, Last Project Evalu

sc = StandardScaler().fit_transform(X)
Z  = linkage(sc, method='ward')
clusters = fcluster(Z, 2, criterion='maxclust')

hr['cluster'] = clusters

median_satisfaction = hr.groupby('cluster')['S'].median().sort_index
cluster_sizes = hr['cluster'].value_counts().sort_index()

print("==== Q4: Median Satisfaction by Segment ====")
print(median_satisfaction)
print("\nCluster Sizes:", cluster_sizes)
```

```
# ================================
# QUESTION 5 — TELCO CLUSTERING (5 clusters)
# ================================

tel_scaled = StandardScaler().fit_transform(tel)
Zt = linkage(tel_scaled, method='ward')
tel_clusters = fcluster(Zt, 5, criterion='maxclust')

tel['groups'] = tel_clusters

agg = tel.groupby('groups').mean()

proptemp = tel.groupby('groups')['Calls'].count()
agg['proportion'] = proptemp / proptemp.sum()

agg = agg.sort_values('proportion', ascending=False)
agg.reset_index(inplace=True)

agg = agg.round(2)

print("\n==== Q5: Aggregated Telco Table ====")
print(agg[['groups','Calls','Intern','Text','Data','Age','proportion


# ================================
# STATEMENT CHECKS FOR QUIZ 5
# ================================

# A: YA uses most data + text
idx_max_text = agg['Text'].idxmax()
idx_max_data = agg['Data'].idxmax()

# C: Heavy User longest calls
idx_max_calls = agg['Calls'].idxmax()

# E: Silver lowest text
idx_min_text = agg['Text'].idxmin()

# D: Did everyone make an intl call?
any_zero_intern = (tel['Intern'] == 0).any()

print("\n==== STATEMENT CHECKS ====")
print("Max Text cluster:", agg.loc[idx_max_text])
print("Max Data cluster:", agg.loc[idx_max_data])
print("Max Calls cluster:", agg.loc[idx_max_calls])
print("Min Text cluster:", agg.loc[idx_min_text])
print("Any Intern = 0 present?:", any_zero_intern)
```

```
print("Any Intern = 0 present?:", any_zero_Intern)
```

```
Please ensure the file upload dialog appears in your browser and yo
Choose Files  no files selected   Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving DATA_2.01_SKU.csv to DATA_2.01_SKU.csv
Saving DATA_2.02_HR.csv to DATA_2.02_HR.csv
Saving DATA_2.03_Telco.csv to DATA_2.03_Telco.csv
Files uploaded successfully (or skipped if already present).
==== Q4: Median Satisfaction by Segment ====
cluster
1    0.41
2    0.64
Name: S, dtype: float64

Cluster Sizes: cluster
1     947
2    1053
Name: count, dtype: int64

==== Q5: Aggregated Telco Table ====
   groups  Calls  Intern    Text  Data    Age  proportion
0       5   3.10    0.52  307.61  2.03  33.58        0.28
1       1   1.71    0.10  509.42  3.73  18.95        0.25
2       3   1.03    0.20   21.89  0.21  60.83        0.20
3       4   5.07    1.01   26.08  2.01  46.69        0.18
4       2   2.30    0.10   21.40  0.52  35.60        0.09

==== STATEMENT CHECKS ====
Max Text cluster: groups          1.00
Calls             1.71
Intern            0.10
Text            509.42
Data              3.73
Age              18.95
proportion        0.25
Name: 1, dtype: float64
Max Data cluster: groups          1.00
Calls             1.71
Intern            0.10
Text            509.42
Data              3.73
Age              18.95
proportion        0.25
Name: 1, dtype: float64
Max Calls cluster: groups         4.00
Calls             5.07
Intern            1.01
Text             26.08
Data              2.01
Age              46.69
proportion        0.18
Name: 3, dtype: float64
Min Text cluster: groups          2.00
```

```
Calls          2.30
Intern         0.10
Text          21.40
Data           0.52
Age           35.60
proportion     0.09
Name: 4, dtype: float64
```