```python
 1  # === kickstarterClean.ipynb ===
 2  #%%
 3  import pandas as pd
 4
 5
 6  #%%
 7  # The path to our CSV file
 8  file = "Resources/KickstarterData.csv"
 9
10  # Read our Kickstarter data into pandas
11  df = pd.read_csv(file)
12  df.head()
13
14
15  #%%
16  # Get a list of all of our columns for easy reference
17  df.columns
18
19
20  #%%
21  # Extract "name", "goal", "pledged", "state", "country", "staff_pick",
22  # "backers_count", and "spotlight"
23  reduced_kickstarter_df = df.loc[:, ["name", "goal", "pledged",
24      "state", "country", "staff_pick", "backers_count", "spotlight"]]
25  reduced_kickstarter_df
26
27
28  #%%
29  # Remove projects that made no money at all
30  reduced_kickstarter_df = reduced_kickstarter_df.loc[(
31      reduced_kickstarter_df["pledged"] > 0)]
32  reduced_kickstarter_df.head()
33
34
35  #%%
36  # Collect only those projects that were hosted in the US
37  hosted_in_us = reduced_kickstarter_df.loc[reduced_kickstarter_df["country"] == "US"]
38  hosted_in_us.head()
39
40
41  #%%
42  # Create a new column that finds the average amount pledged to a project
43  hosted_in_us["average_donation"] = hosted_in_us['pledged'] /
44      hosted_in_us['backers_count']
45
46
47  #%%
48  # First convert "average_donation", "goal", and "pledged" columns to float
49  # Then Format to go to two decimal places, include a dollar sign, and use comma notation
50
51  hosted_in_us["average_donation"] = hosted_in_us["average_donation"].astype(float).map(
52      "${:,.2f}".format)
53  hosted_in_us["goal"] = hosted_in_us["goal"].astype(float).map("${:,.2f}".format)
54  hosted_in_us["pledged"] = hosted_in_us["pledged"].astype(float).map("${:,.2f}".format)
55
```

```
56
57   #%%
58   hosted_in_us.head()
59
60
61   #%%
62   # Calculate the total number of backers for all US projects
63   hosted_in_us["backers_count"].sum()
64
65
66   #%%
67   # Calculate the average number of backers for all US projects
68   hosted_in_us["backers_count"].mean()
69
70
71   #%%
72   # Collect only those US campaigns that have been picked as a "Staff Pick"
73   picked_by_staff = hosted_in_us.loc[hosted_in_us["staff_pick"] == True]
74   picked_by_staff
75
76
77   #%%
78   # Group by the state of the campaigns and see if staff picks matter (Seems to matter quite
…    a bit)
79   state_groups = picked_by_staff.groupby("state")
80   state_groups["name"].count()
81
82
83
```