

Exploratory Data Analysis - AltMetric research paper

In this file we are going to perform an exploratory data analysis for AltMetric data. Please have in mind the methodology in use, i.e. CRISP-DM have some stages such as data retention, data cleaning and etc. which is necessary for EDA, except data retention which is done via AltMetric platform, other necessary steps will be done in this notebook.

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from sklearn import preprocessing
pd.options.display.max_columns = 200
sns.set_theme(style='darkgrid')
```

Data Gathering

In this section we are going to import the gathered data in the code for further analysis.

```
In [ ]: atu = pd.read_csv('data/allameh.csv')
atu['University'] = "Allameh Tabataba'i University"

aut = pd.read_csv('data/amirkabir.csv')
aut['University'] = "Amir Kabir University"

sbu = pd.read_csv('data/beheshti.csv')
sbu['University'] = "Shahid Beheshti University"

fum = pd.read_csv('data/ferdowsi.csv')
fum['University'] = "Ferdowsi University of Mashhad"

ugui = pd.read_csv('data/guilan.csv')
ugui['University'] = "University of Guilan"

ihu = pd.read_csv('data/imamhosein.csv')
ihu['University'] = "Imam Hossein University"

uisf = pd.read_csv('data/isfahan.csv')
uisf['University'] = "University of Isfahan"

iut = pd.read_csv('data/iut.csv')
iut['University'] = "Isfahan University of Technology"

knu = pd.read_csv('data/knu.csv')
knu['University'] = "K. N. Toosi University of Technology"

sut = pd.read_csv('data/sharif.csv')
sut['University'] = "Sharif University of Technology"

ushi = pd.read_csv('data/shiraz.csv')
ushi['University'] = "University of Shiraz"

iust = pd.read_csv('data/stu.csv')
iust['University'] = "Iran University of Science and Technology"

utab = pd.read_csv('data/tabriz.csv')
utab['University'] = "University of Tabriz"

tmu = pd.read_csv('data/tarbiatmodares.csv')
tmu['University'] = "Tarbiat Modares University"

uteh = pd.read_csv('data/ut.csv')
uteh['University'] = "University of Tehran"
```

Data Transformation & Feature Engineering

In the next section we are going to transform imported data to our needs and some feature generations are in order, i.e. label encoding two categorical variables, renaming features for standard pandas practice and etc. Another important thing we are doing in the cell below, is that we will split the column "Subjects_FoR" to main categories, to the length of the record with the highest category items.

On the context of categories, it's worth to mention some explanation regarding the matter. Based on the retrieved data, the Feature "Subjects_FoR" contains all categories for each article, these categories are based on standard system of categorization of AltMetric, the categories which starts with a 2 digit number, are major categories and those one which starts with 4 digit numbers are the sub categories of their corresponding major category. For this scientific endeavor we will only focus on the major categories.

As we explored this feature, we found out that the article with the most major categories actually have 7 major categories, thus we create 7 columns one for each major categories, obviously the categories with greater number are more probable to have "No Category" as value.

```
In [ ]: df = pd.concat([atu, aut, sbu, fum, ugui, ihu, uisf, iut,
                      knu, sut, ushi, iust, utab, tmu, uteh])

df.drop(['Authors at my Institution', 'Departments', 'Journal ISSNs', 'Sustainable
         df.columns = df.columns.str.replace(' ', '_')
         df = df.reset_index(drop=True)

df['Subjects_(FoR)'].fillna('00 No Category', inplace=True)
df['Journal/Collection_Title'].fillna('No Title', inplace=True)
main_lst = []
for record in df['Subjects_(FoR)'].str.split('; '):
    temp_lst = []
    for element in record:
        if len(element.split(' '))[0] == 2:
            temp_lst.append(element)
    main_lst.append(temp_lst)
temp_df = pd.DataFrame(main_lst, columns=[

    'Category_1',
    'Category_2',
    'Category_3',
    'Category_4',
    'Category_5',
    'Category_6',
    'Category_7',
])
temp_df.fillna('00 No Category', inplace=True)
df = pd.concat([df, temp_df], axis=1)

label_encoder = preprocessing.LabelEncoder()
df['Journal/Collection_Title_LE'] = label_encoder.fit_transform(df['Journal/Collect
df['Output_Type_LE'] = label_encoder.fit_transform(df['Output_Type'])
df['OA_Status_LE'] = label_encoder.fit_transform(df['OA_Status'])
df['OA_Type_LE'] = label_encoder.fit_transform(df['OA_Type'])
df['Publisher_Names_LE'] = label_encoder.fit_transform(df['Publisher_Names'])
df['University_LE'] = label_encoder.fit_transform(df['University'])
df['Category_1_LE'] = label_encoder.fit_transform(df['Category_1'])
df['Category_2_LE'] = label_encoder.fit_transform(df['Category_2'])
df['Category_3_LE'] = label_encoder.fit_transform(df['Category_3'])
df['Category_4_LE'] = label_encoder.fit_transform(df['Category_4'])
df['Category_5_LE'] = label_encoder.fit_transform(df['Category_5'])
df['Category_6_LE'] = label_encoder.fit_transform(df['Category_6'])
df['Category_7_LE'] = label_encoder.fit_transform(df['Category_7'])
df['Publication_Date'] = pd.to_datetime(df['Publication_Date'])
```

```
In [ ]: df['OA_Type'].value_counts()
```

```
Out[ ]: closed    48155
         gold      11523
         green     9690
         bronze    2656
         hybrid    1985
Name: OA_Type, dtype: int64
```

In the table below you can see the top 5 and bottom 5 records in the main dataframe. In this dataframe we have done the feature engineering generation phase.

This dataframe consists of some features that we will going to describe each one of them:

- **Altmetric_Attention_Score:** This is the main feature of the dataset, this metric is the attention score for each article, this feature is seems like the aggregation of multiple features which we will discuss further in this section.
- **Title:** The title of article.
- **Journal/Collection_Title:** The journal or collection the article have published in. Types other than "Article" will not have title since they are not published in the journal or collection.
- **Output_Type:** This is a categorical feature, 4 classes are used in this feature: 1. Article, 2. Chapter, 3. Book, 4. News. The classes are self explanatory.
- **OA_Status:** Whether that article is published under open access license or not.
- **OA_Type:** A categorical feature determining the type of open access each class is discussed:
 1. Closed: This article is not published in open access journal.
 2. Gold: The most open and least restrictive type of open access.
 3. Green: This is when the accepted article is first deposited into a subject-based repository or an institution's repository, which then often specifies how the article may be used.
 4. Bronze: This is not fully open access because although the article will be freely available, the types of open access journals that offer this kind of service have no open license.
 5. Hybrid: This is one of those types of open access where a subscription journal offers open access, where the processing fee is paid for individual articles. Although the processing fees may be higher than that of a regular open access journal, it may be worth the effort if your article fits the journal's aims and scope perfectly.
- **Subjects_(FoR):** The categories used for each record, these categories are recorded as string and concatenated together.
- **Affiliations_(GRID):** The affiliations of the authors.
- **Publication_Date:** The date that record have been published.
- **News_mentions:** The number of mentions that the corresponding record had in the *News*.
- **Blog_mentions:** The number of mentions that the corresponding record had in the *Blog*.
- **Policy_mentions:** The number of mentions that record used in the *Policies*.
- **Patent_mentions:** The number of times that the corresponding record was mentioned in the *Patents*.
- **Twitter_mentions:** The number of mentions that the corresponding record had in the *Twitter*.
- **Peer_review_mentions:** The number of mentions that the corresponding record had in

the *Peer Review*.

- **Weibo_mentions:** The number of mentions that the corresponding record had in the *Weibo*.
- **Facebook_mentions:** The number of mentions that the corresponding record had in the *Facebook*.
- **Wikipedia_mentions:** The number of mentions that the corresponding record had in the *Wikipedia*.
- **Google+_mentions:** The number of mentions that the corresponding record had in the *Google+*.
- **Linkedin_mentions:** The number of mentions that the corresponding record had in the *LinkedIn*.
- **Reddit_mentions:** The number of mentions that the corresponding record had in the *Reddit*.
- **Pinterest_mentions:** The number of mentions that the corresponding record had in the *Pinterest*.
- **F1000_mentions:** The number of mentions that the corresponding record had in the *F1000*. F1000 is a open research publisher for scientists, scholars, and clinical researchers.
- **Q&A_mentions:** The number of mentions that the corresponding record had in the *Q&A Services*.
- **Video_mentions:** The number of mentions that the corresponding record had in the *Video*.
- **Number_of_Mendeley_readers:** The number of Mendeley users that have added a particular document to a *Mendeley* library.
- **Number_of_Dimensions_citations:** Dimensions extracts references between publications either from existing databases (such as CrossRef, PubMed Central or OpenCitations data), or directly from the full text record provided by the content publisher. Reference extraction is not limited to journal items, but also includes citations from and to books, conference proceedings and pre-prints.
- **Publisher_Names:** The name of publisher which have published the record.
- **University:** The name of university that this record related to.
- **Category_1:** The *First* major category used in this record.
- **Category_2:** The *Second* major category used in this record.
- **Category_3:** The *Third* major category used in this record.
- **Category_4:** The *Fourth* major category used in this record.
- **Category_5:** The *Fifth* major category used in this record.
- **Category_6:** The *Sixth* major category used in this record.
- **Category_7:** The *Seventh* major category used in this record.

Features with LE at their end is label encoded version of a corresponding feature.

In []: df

Out[]:	Altmetric_Attention_Score	Title	Journal/Collection_Title	Output_Type	OA_Status
0	1629	COVID-19 and male reproductive function: a pro...	Reproduction	Article	True
1	906	The effects of three different exercise modalit...	Reproduction	Article	True
2	239	Fear, Loss, Social Isolation, and Incomplete G...	Basic And Clinical Neuroscience	Article	True
3	175	Foundations of Social Policy and Welfare in Islam	No Title	Chapter	False
4	102	Laughter yoga versus group exercise program in...	International Journal of Geriatric Psychiatry	Article	True
...
74004	0	AN AHP MODEL FOR CROP PLANNING WITHIN IRRIGATI...	Irrigation & Drainage	Article	False
74005	0	GC-MS Determination of PAHs in Fish Samples Fo...	Chromatographia	Article	False
74006	0	Interface thermal resistance and thermal recti...	Applied Physics Letters	Article	False
74007	0	The systematic importance of anatomical data i...	Botanical Journal of the Linnean Society	Article	True

	Altmetric_Attention_Score	Title	Journal/Collection_Title	Output_Type	OA_Status
74008	0	Pollen morphology of the genus Gagea (Liliacea...		Flora Article	False

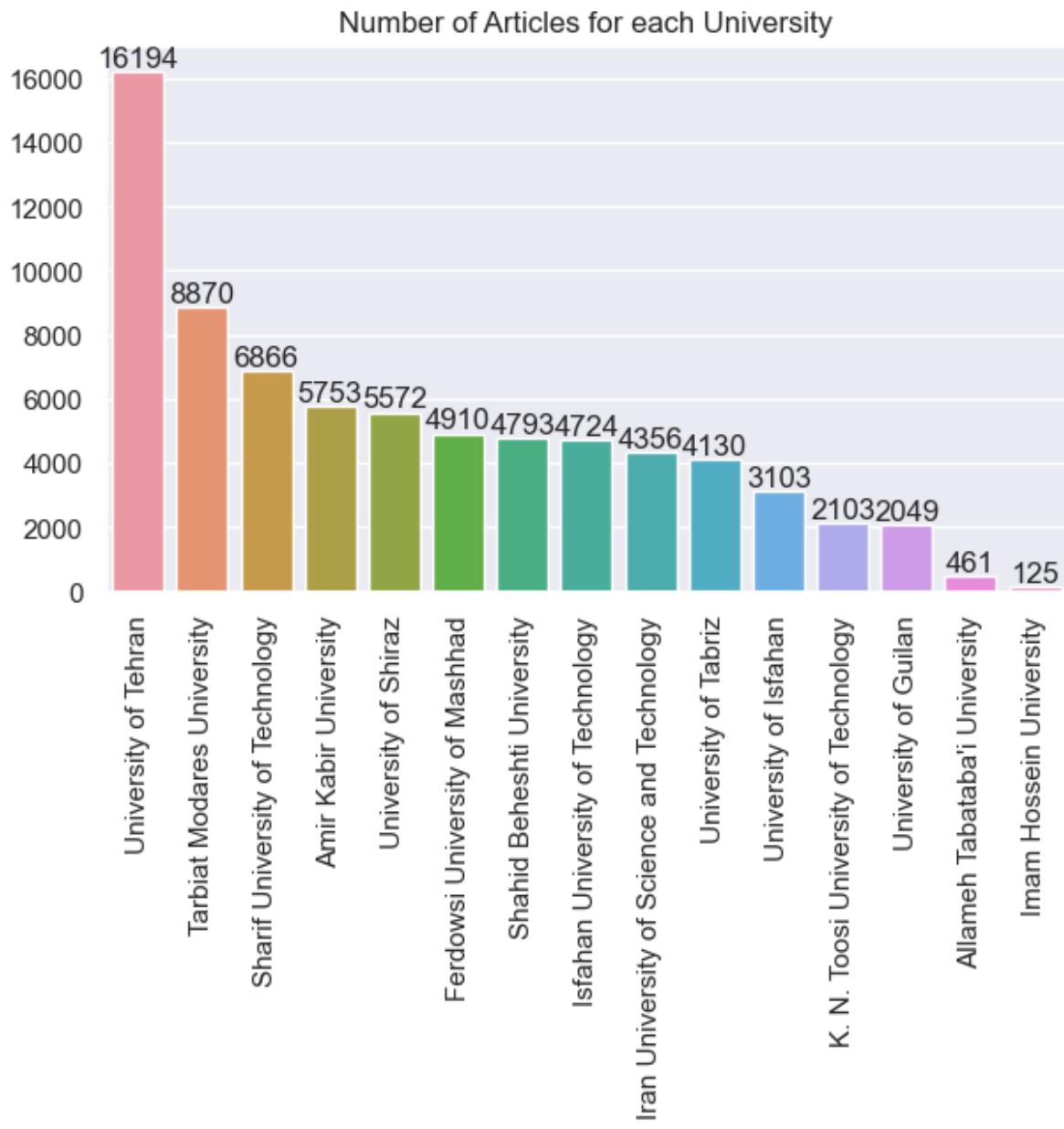
74009 rows × 49 columns

Data Exploration

First thing first, we must compare the number of article for each university. In the plot below, you can see the number of articles per university, sorted in a descending order.

As you can see the top university on the number of research output is **University of Tehran** with more than 16,000 articles. In the second and third place are **Tarbiat Modares University** and **Sharif University of Technology**. The lowest three universities on the context of research output are: **Imam Hossein University**, **Allameh Tabataba'i University** and **University of Guilan**.

```
In [ ]: x = list(df['University'].value_counts().index)
y = list(df['University'].value_counts())
plt.figure(figsize=(7, 4))
ax = sns.barplot(data=df, x=x, y=y)
ax.bar_label(ax.containers[0])
plt.title('Number of Articles for each University')
plt.xticks(rotation=90)
plt.show()
```



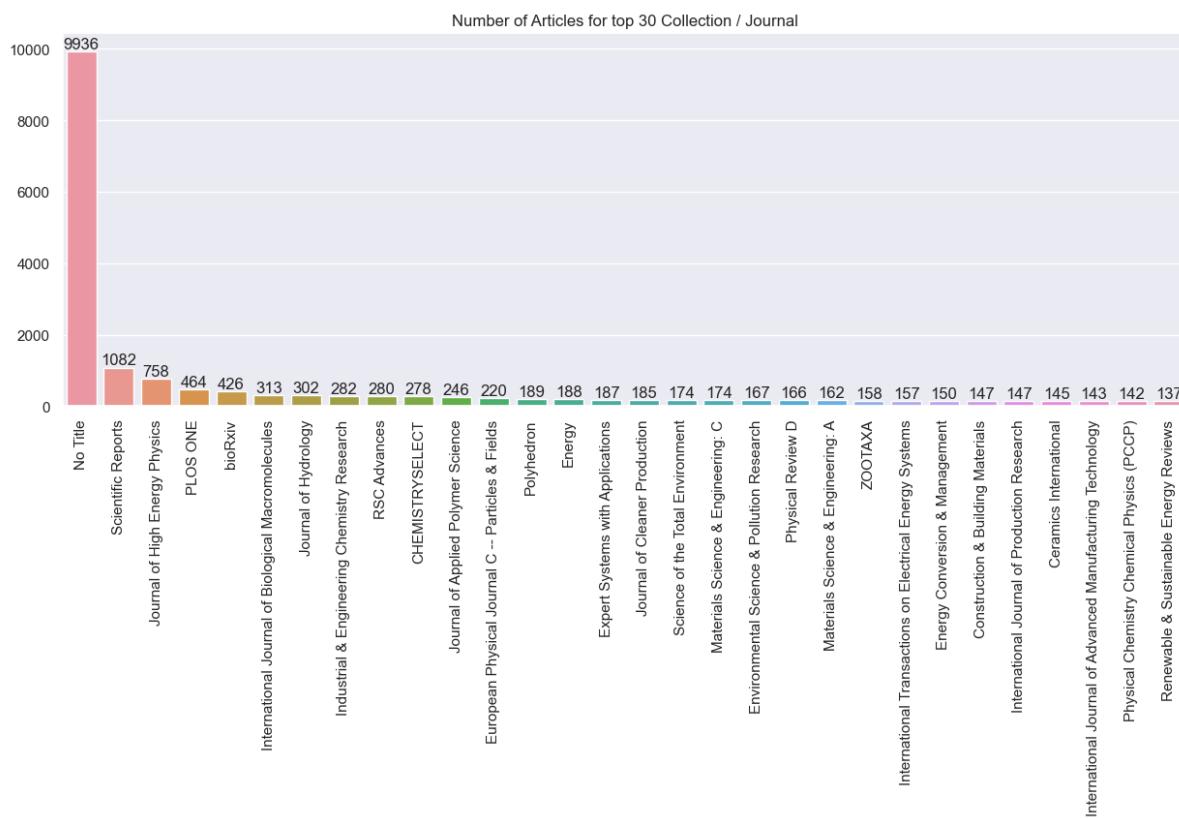
In the plot below, you can see the top 30 journal or collection in which universities have published their research output.

The first place is **No Title**, this fact is due to the some of research outputs are not articles published in journals, they are books, chapters and news, this kind of articles have no corresponding journal title and this feature is due to this fact.

The journal with the most hosted articles are **Scientific Reports** which is a open access scientific journal under Nature and was the 5th most cited journal in the world.

On the second and third place we have **Journal of High Energy Physics** and **PLOS ONE**.

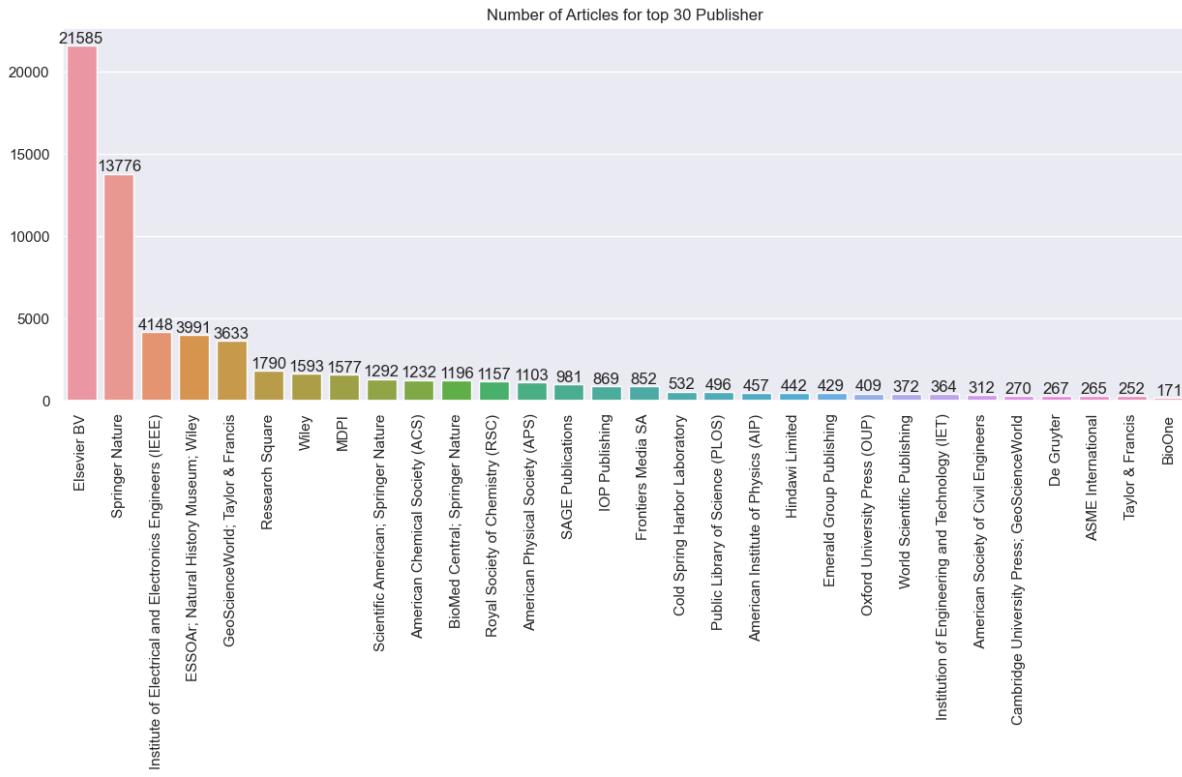
```
In [ ]: x = list(df['Journal/Collection_Title'].value_counts().head(30).index)
y = list(df['Journal/Collection_Title'].value_counts().head(30))
plt.figure(figsize=(15, 5))
ax = sns.barplot(x=x, y=y)
ax.bar_label(ax.containers[0])
plt.title('Number of Articles for top 30 Collection / Journal')
plt.xticks(rotation=90)
plt.show()
```



In the plot below, you can see the top 30 Publisher in which universities have published their research output.

The publisher with the most articles of these universities are **Elsevier BV** with more than 21,000 and in the second place with more than 13,000 articles is **Springer Nature**.

```
In [ ]: x = list(df['Publisher_Names'].value_counts().head(30).index)
y = list(df['Publisher_Names'].value_counts().head(30))
plt.figure(figsize=(15, 5))
ax = sns.barplot(x=x, y=y)
ax.bar_label(ax.containers[0])
plt.title('Number of Articles for top 30 Publisher')
plt.xticks(rotation=90)
plt.show()
```



Categories Exploration

In the following section we will explore the categories quantity status. The plot in the following cell, is the barplot of the 7 category columns.

- More than 16,000 articles published had **Engineering** as first categories, **Medical and Health Sciences** and **Chemical Sciences** is in the second and third place.
- It's Interesting to mention that the Engineering category used as first category is coded 09 but the Engineering category used in other than first places are coded 40.
- More than 17,000 articles had **Engineering** as second category, **Biomedical and Clinical Sciences** was the second most used category as second category.
- Majority of articles (65.4%) had just two categories.
- The similar trend of having the **Engineering** as the most used category is tracked among the other places of category, it is used more than 5,900 times as third category.
- Around 9.6% of articles had more than three major categories.
- **Engineering** is the most used major category as 4th category also.
- 1.8% of articles have 5 and more major categories, 0.3% of articles have 6 and more major categories and just 3 articles have 7 major categories, which all of them had **Mathematical Sciences** as their 7th major category.

```
In [ ]: fig, axes = plt.subplots(7, 1, figsize=(20, 50))
x = [
    list(df['Category_1'].value_counts().index),
    list(df['Category_2'].value_counts().index),
    list(df['Category_3'].value_counts().index),
    list(df['Category_4'].value_counts().index),
    list(df['Category_5'].value_counts().index),
    list(df['Category_6'].value_counts().index),
    list(df['Category_7'].value_counts().index),
]

y = [
    list(df['Category_1'].value_counts()),
    list(df['Category_2'].value_counts()),
    list(df['Category_3'].value_counts()),
    list(df['Category_4'].value_counts()),
    list(df['Category_5'].value_counts()),
    list(df['Category_6'].value_counts()),
    list(df['Category_7'].value_counts()),
]

sns.barplot(ax=axes[0], x=x[0], y=y[0])
axes[0].set_title('Number of First Categories used in Articles')
axes[0].bar_label(axes[0].containers[0])
axes[0].tick_params(labelrotation=90)

sns.barplot(ax=axes[1], x=x[1], y=y[1])
axes[1].set_title('Number of Second Categories used in Articles')
axes[1].bar_label(axes[1].containers[0])
axes[1].tick_params(labelrotation=90)

sns.barplot(ax=axes[2], x=x[2], y=y[2])
axes[2].set_title('Number of Third Categories used in Articles')
axes[2].bar_label(axes[2].containers[0])
axes[2].tick_params(labelrotation=90)

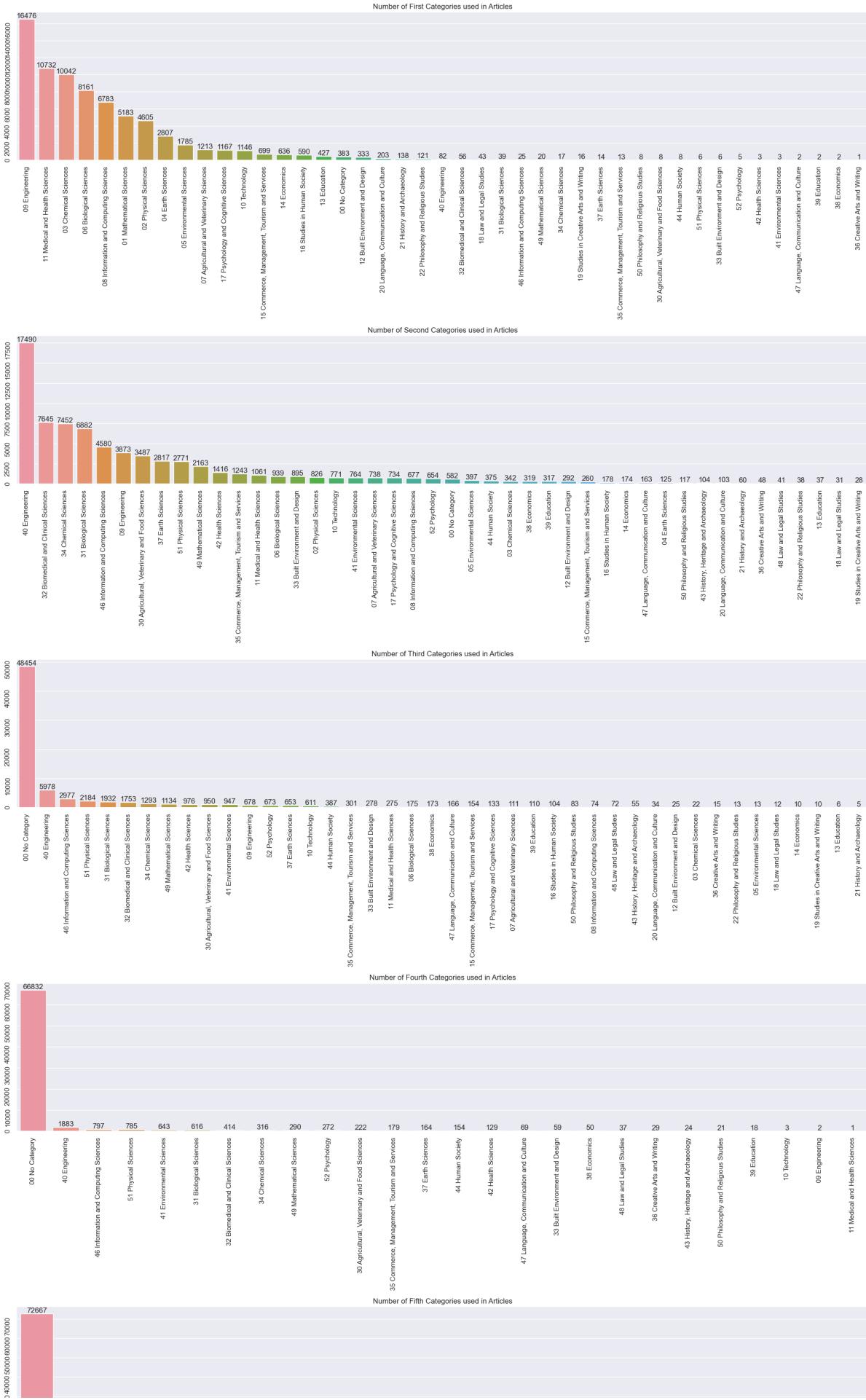
sns.barplot(ax=axes[3], x=x[3], y=y[3])
axes[3].set_title('Number of Fourth Categories used in Articles')
axes[3].bar_label(axes[3].containers[0])
axes[3].tick_params(labelrotation=90)

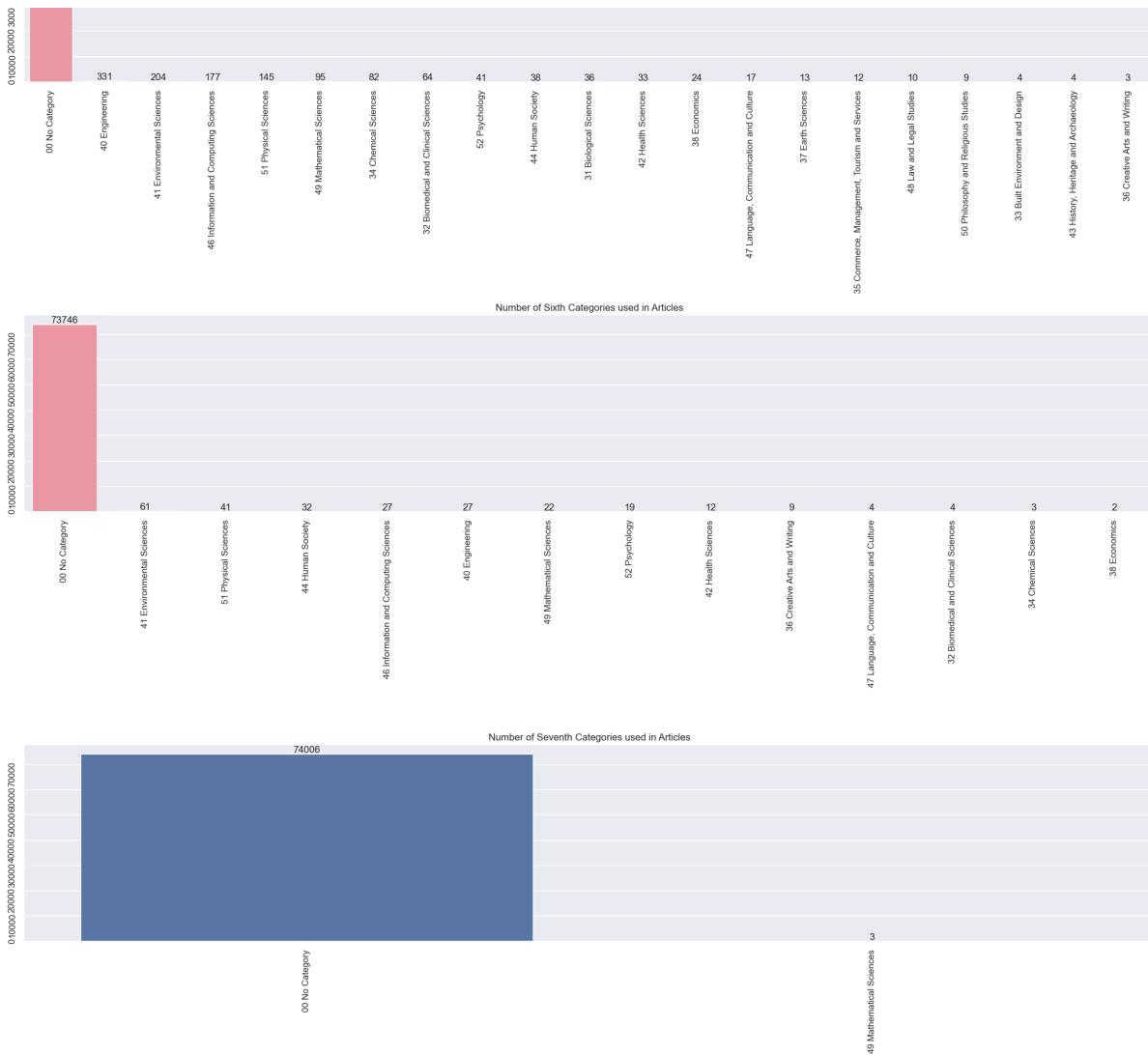
sns.barplot(ax=axes[4], x=x[4], y=y[4])
axes[4].set_title('Number of Fifth Categories used in Articles')
axes[4].bar_label(axes[4].containers[0])
axes[4].tick_params(labelrotation=90)

sns.barplot(ax=axes[5], x=x[5], y=y[5])
axes[5].set_title('Number of Sixth Categories used in Articles')
axes[5].bar_label(axes[5].containers[0])
axes[5].tick_params(labelrotation=90)

sns.barplot(ax=axes[6], x=x[6], y=y[6])
axes[6].set_title('Number of Seventh Categories used in Articles')
axes[6].bar_label(axes[6].containers[0])
axes[6].tick_params(labelrotation=90)

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





In the Cells Below we will explore the 7 Categories for each University.

Allameh Tabataba'i University top 7 Categories

- 25.1% of published articles had **Medical and Health Sciences** as first major category.
- **Engineering** is the 8th most used first category in the articles of the ATU.
- This might be an indicator toward the Engineering articles, meaning the attention score by altmetric might be biased towards these type of articles, which can be rooted to the interest of media and news agency towards engineering breakthroughs.
- Majority of articles published by ATU (63.1%) had two and less major categories.
- No articles published by ATU had 7 major categories.

```
In [ ]: fig, axes = plt.subplots(3, 3, figsize=(20, 20))
x = [
    list(df[df['University'] == "Allameh Tabataba'i University"]['Category_1'].value),
    list(df[df['University'] == "Allameh Tabataba'i University"]['Category_2'].value),
    list(df[df['University'] == "Allameh Tabataba'i University"]['Category_3'].value),
    list(df[df['University'] == "Allameh Tabataba'i University"]['Category_4'].value),
    list(df[df['University'] == "Allameh Tabataba'i University"]['Category_5'].value),
    list(df[df['University'] == "Allameh Tabataba'i University"]['Category_6'].value),
    list(df[df['University'] == "Allameh Tabataba'i University"]['Category_7'].value)
]
y = [
    list(df[df['University'] == "Allameh Tabataba'i University"]['Category_1'].value),
    list(df[df['University'] == "Allameh Tabataba'i University"]['Category_2'].value),
    list(df[df['University'] == "Allameh Tabataba'i University"]['Category_3'].value),
    list(df[df['University'] == "Allameh Tabataba'i University"]['Category_4'].value),
    list(df[df['University'] == "Allameh Tabataba'i University"]['Category_5'].value),
    list(df[df['University'] == "Allameh Tabataba'i University"]['Category_6'].value),
    list(df[df['University'] == "Allameh Tabataba'i University"]['Category_7'].value)
]
sns.barplot(ax=axes[0, 0], x=x[0], y=y[0])
axes[0, 0].set_title("Number of First Categories used in Articles of Allameh Tabataba'i University")
axes[0, 0].bar_label(axes[0, 0].containers[0])
axes[0, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 1], x=x[1], y=y[1])
axes[0, 1].set_title("Number of Second Categories used in Articles of Allameh Tabataba'i University")
axes[0, 1].bar_label(axes[0, 1].containers[0])
axes[0, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 2], x=x[2], y=y[2])
axes[0, 2].set_title("Number of Third Categories used in Articles of Allameh Tabataba'i University")
axes[0, 2].bar_label(axes[0, 2].containers[0])
axes[0, 2].tick_params(labelrotation=90)

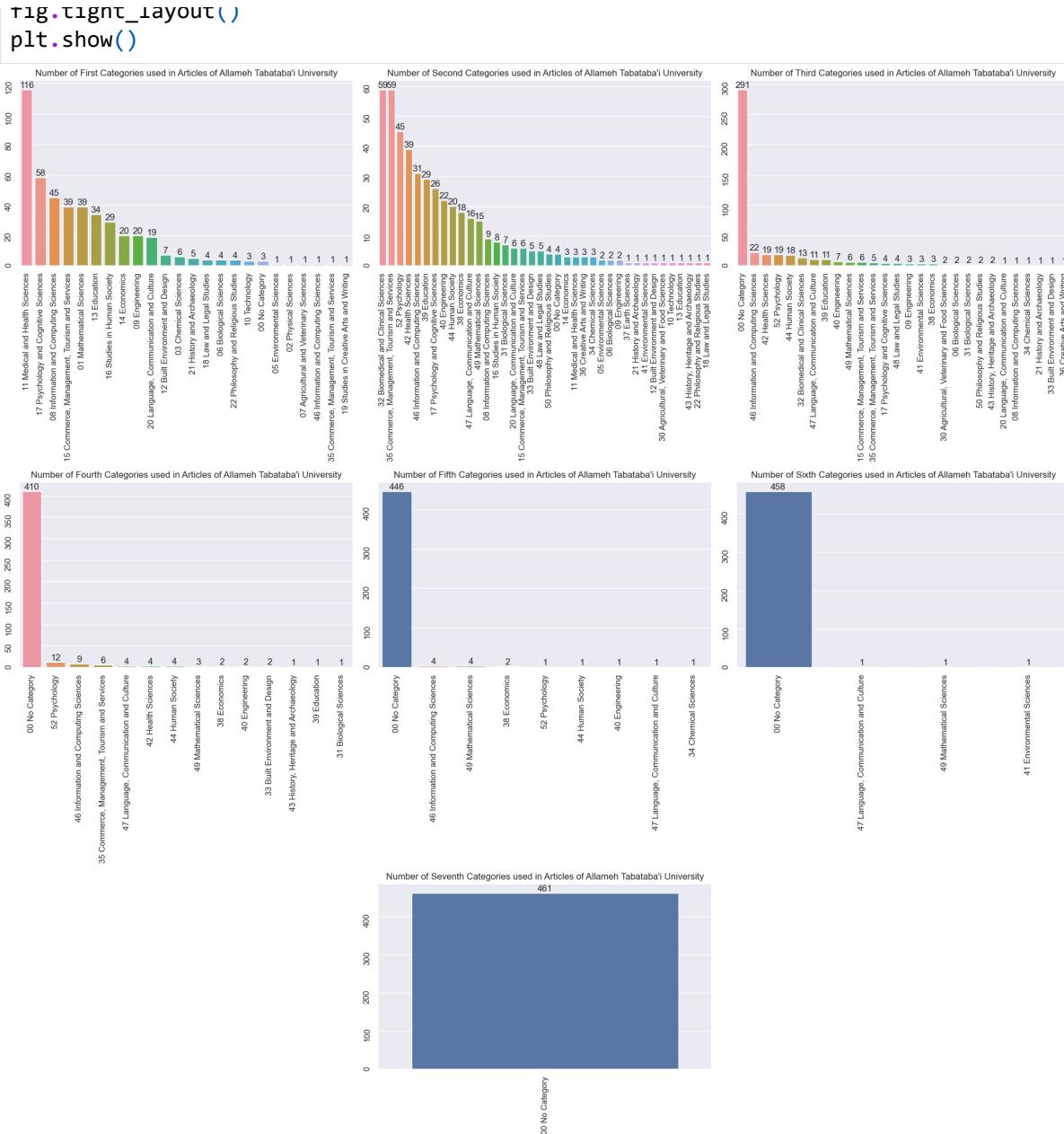
sns.barplot(ax=axes[1, 0], x=x[3], y=y[3])
axes[1, 0].set_title("Number of Fourth Categories used in Articles of Allameh Tabataba'i University")
axes[1, 0].bar_label(axes[1, 0].containers[0])
axes[1, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 1], x=x[4], y=y[4])
axes[1, 1].set_title("Number of Fifth Categories used in Articles of Allameh Tabataba'i University")
axes[1, 1].bar_label(axes[1, 1].containers[0])
axes[1, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 2], x=x[5], y=y[5])
axes[1, 2].set_title("Number of Sixth Categories used in Articles of Allameh Tabataba'i University")
axes[1, 2].bar_label(axes[1, 2].containers[0])
axes[1, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[2, 1], x=x[6], y=y[6])
axes[2, 1].set_title("Number of Seventh Categories used in Articles of Allameh Tabataba'i University")
axes[2, 1].bar_label(axes[2, 1].containers[0])
axes[2, 1].tick_params(labelrotation=90)

fig.delaxes(axes[2, 0])
fig.delaxes(axes[2, 2])
```



Amir Kabir University Top 7 Categories

- 38.3% of published articles of AUT have **Engineering** as their first category, **Information and Computing Sciences** and **Chemical Sciences** are in the second and third place.
- 42.4% of published articles of AUT have **Engineering** as their second category and like the first item, **Information and Computing Sciences** and **Chemical Sciences** are in second and third place.
- The majority of published articles of AUT (65.6%) have one or two major categories.

```
In [ ]: fig, axes = plt.subplots(3, 3, figsize=(20, 20))
x = [
    list(df[df['University'] == "Amir Kabir University"]['Category_1'].value_counts),
    list(df[df['University'] == "Amir Kabir University"]['Category_2'].value_counts),
    list(df[df['University'] == "Amir Kabir University"]['Category_3'].value_counts),
    list(df[df['University'] == "Amir Kabir University"]['Category_4'].value_counts),
    list(df[df['University'] == "Amir Kabir University"]['Category_5'].value_counts),
    list(df[df['University'] == "Amir Kabir University"]['Category_6'].value_counts),
    list(df[df['University'] == "Amir Kabir University"]['Category_7'].value_counts)
]
y = [
    list(df[df['University'] == "Amir Kabir University"]['Category_1'].value_counts),
    list(df[df['University'] == "Amir Kabir University"]['Category_2'].value_counts),
    list(df[df['University'] == "Amir Kabir University"]['Category_3'].value_counts),
    list(df[df['University'] == "Amir Kabir University"]['Category_4'].value_counts),
    list(df[df['University'] == "Amir Kabir University"]['Category_5'].value_counts),
    list(df[df['University'] == "Amir Kabir University"]['Category_6'].value_counts),
    list(df[df['University'] == "Amir Kabir University"]['Category_7'].value_counts)
]
sns.barplot(ax=axes[0, 0], x=x[0], y=y[0])
axes[0, 0].set_title("Number of First Categories used in Articles of Amir Kabir Uni")
axes[0, 0].bar_label(axes[0, 0].containers[0], rotation=90)
axes[0, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 1], x=x[1], y=y[1])
axes[0, 1].set_title("Number of Second Categories used in Articles of Amir Kabir Un")
axes[0, 1].bar_label(axes[0, 1].containers[0], rotation=90)
axes[0, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 2], x=x[2], y=y[2])
axes[0, 2].set_title("Number of Third Categories used in Articles of Amir Kabir Uni")
axes[0, 2].bar_label(axes[0, 2].containers[0], rotation=90)
axes[0, 2].tick_params(labelrotation=90)

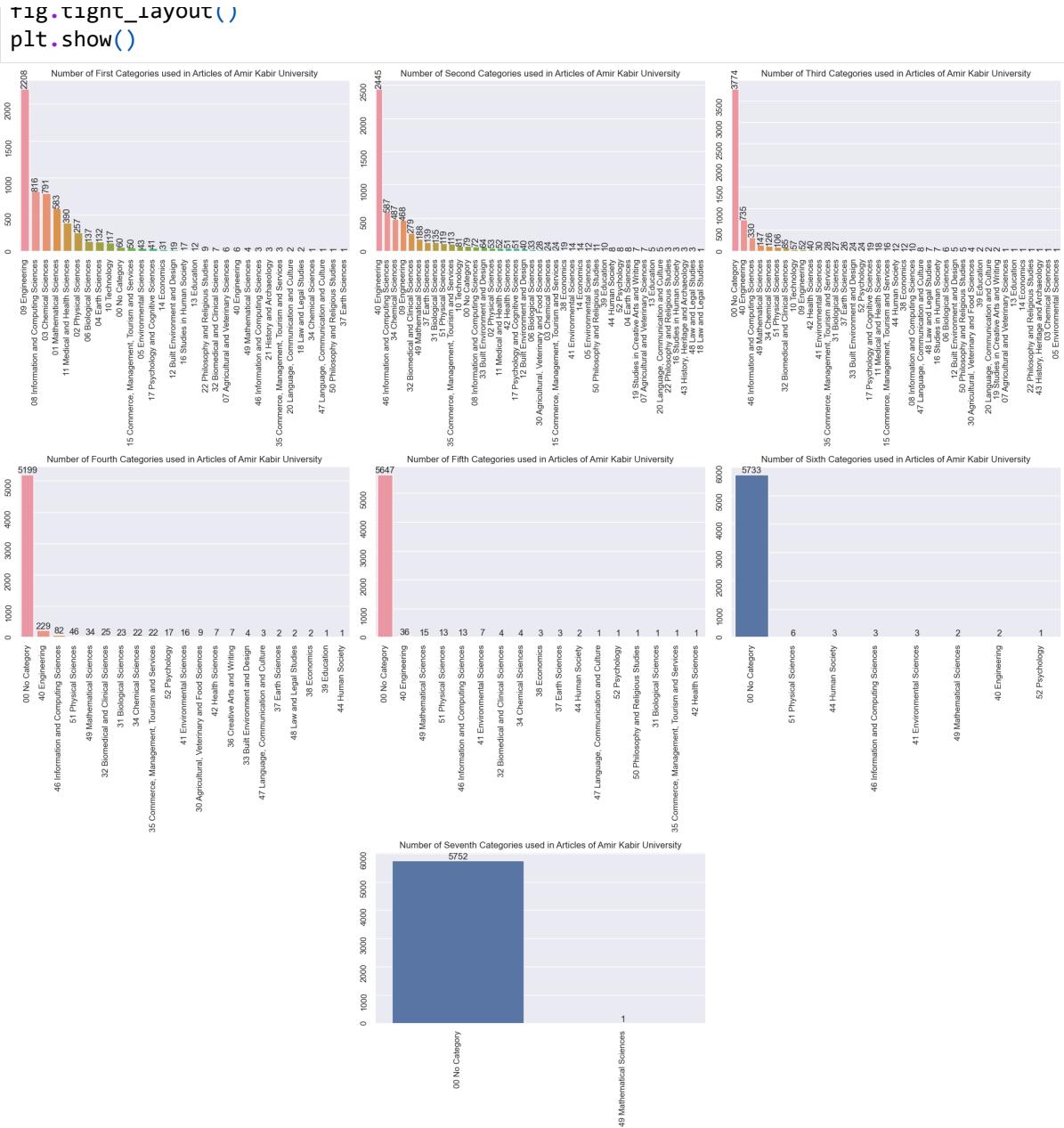
sns.barplot(ax=axes[1, 0], x=x[3], y=y[3])
axes[1, 0].set_title("Number of Fourth Categories used in Articles of Amir Kabir Un")
axes[1, 0].bar_label(axes[1, 0].containers[0])
axes[1, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 1], x=x[4], y=y[4])
axes[1, 1].set_title("Number of Fifth Categories used in Articles of Amir Kabir Uni")
axes[1, 1].bar_label(axes[1, 1].containers[0])
axes[1, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 2], x=x[5], y=y[5])
axes[1, 2].set_title("Number of Sixth Categories used in Articles of Amir Kabir Uni")
axes[1, 2].bar_label(axes[1, 2].containers[0])
axes[1, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[2, 1], x=x[6], y=y[6])
axes[2, 1].set_title("Number of Seventh Categories used in Articles of Amir Kabir U")
axes[2, 1].bar_label(axes[2, 1].containers[0])
axes[2, 1].tick_params(labelrotation=90)

fig.delaxes(axes[2, 0])
fig.delaxes(axes[2, 2])
```



Shahid Beheshti University top 7 Categories

- The distribution of categories in SBU is much more smoother, in other words, SBU is academically more engaged in different subjects.
- The top category of published articles in SBU is **Medical and Health Sciences**. In the second and third place are **Chemical Sciences** and **Engineering**.
- The majority of published articles in SBU (65.5%) has less than two or one major categories.

```
In [ ]: fig, axes = plt.subplots(3, 3, figsize=(20, 20))
x = [
    list(df[df['University'] == "Shahid Beheshti University"]['Category_1'].value_c),
    list(df[df['University'] == "Shahid Beheshti University"]['Category_2'].value_c),
    list(df[df['University'] == "Shahid Beheshti University"]['Category_3'].value_c),
    list(df[df['University'] == "Shahid Beheshti University"]['Category_4'].value_c),
    list(df[df['University'] == "Shahid Beheshti University"]['Category_5'].value_c),
    list(df[df['University'] == "Shahid Beheshti University"]['Category_6'].value_c),
    list(df[df['University'] == "Shahid Beheshti University"]['Category_7'].value_c)
]
y = [
    list(df[df['University'] == "Shahid Beheshti University"]['Category_1'].value_c),
    list(df[df['University'] == "Shahid Beheshti University"]['Category_2'].value_c),
    list(df[df['University'] == "Shahid Beheshti University"]['Category_3'].value_c),
    list(df[df['University'] == "Shahid Beheshti University"]['Category_4'].value_c),
    list(df[df['University'] == "Shahid Beheshti University"]['Category_5'].value_c),
    list(df[df['University'] == "Shahid Beheshti University"]['Category_6'].value_c),
    list(df[df['University'] == "Shahid Beheshti University"]['Category_7'].value_c)
]
sns.barplot(ax=axes[0, 0], x=x[0], y=y[0])
axes[0, 0].set_title("Number of First Categories used in Articles of Shahid Beheshti University")
axes[0, 0].bar_label(axes[0, 0].containers[0], rotation=90)
axes[0, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 1], x=x[1], y=y[1])
axes[0, 1].set_title("Number of Second Categories used in Articles of Shahid Beheshti University")
axes[0, 1].bar_label(axes[0, 1].containers[0], rotation=90)
axes[0, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 2], x=x[2], y=y[2])
axes[0, 2].set_title("Number of Third Categories used in Articles of Shahid Beheshti University")
axes[0, 2].bar_label(axes[0, 2].containers[0], rotation=90)
axes[0, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 0], x=x[3], y=y[3])
axes[1, 0].set_title("Number of Fourth Categories used in Articles of Shahid Beheshti University")
axes[1, 0].bar_label(axes[1, 0].containers[0])
axes[1, 0].tick_params(labelrotation=90)

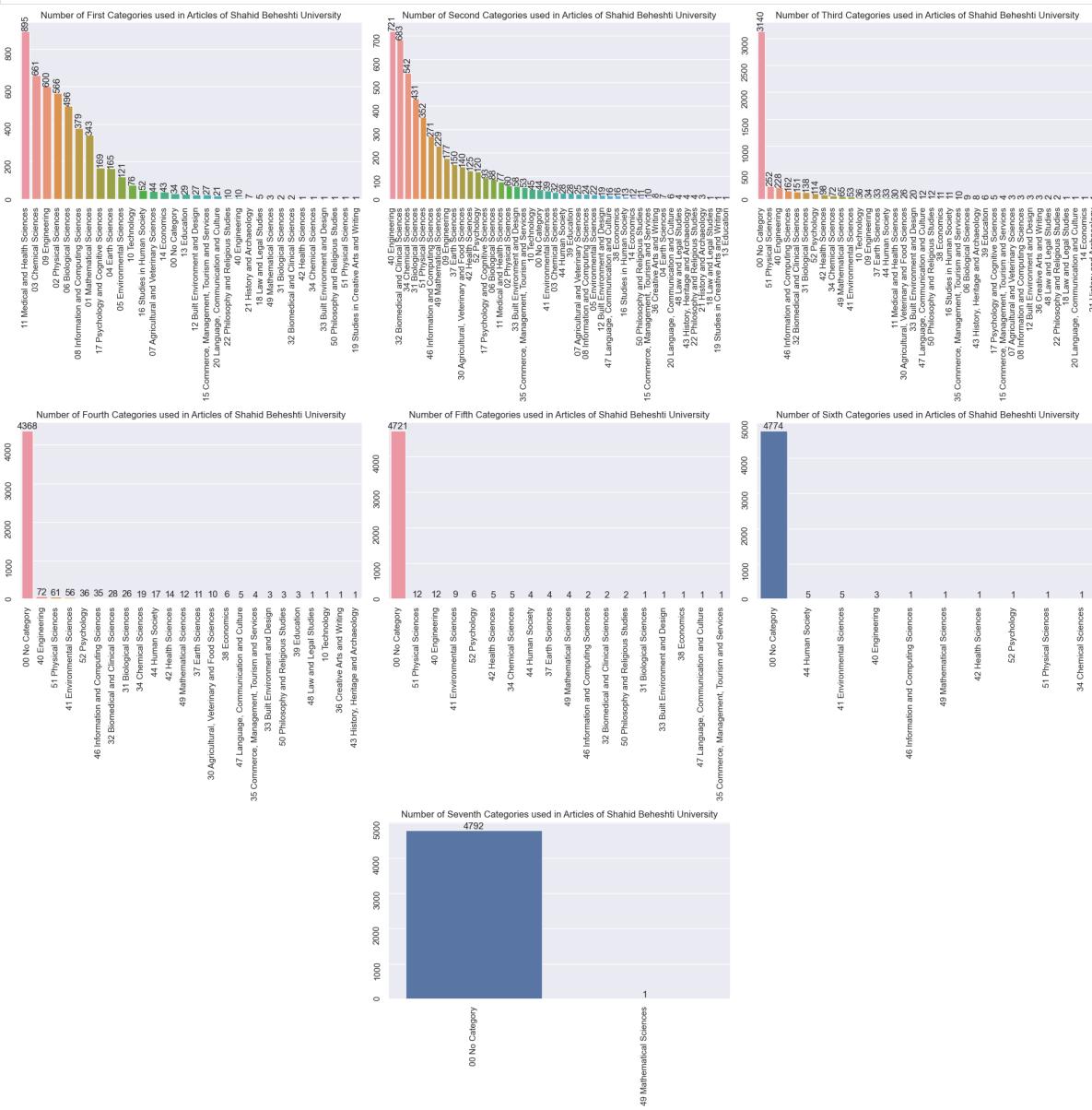
sns.barplot(ax=axes[1, 1], x=x[4], y=y[4])
axes[1, 1].set_title("Number of Fifth Categories used in Articles of Shahid Beheshti University")
axes[1, 1].bar_label(axes[1, 1].containers[0])
axes[1, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 2], x=x[5], y=y[5])
axes[1, 2].set_title("Number of Sixth Categories used in Articles of Shahid Beheshti University")
axes[1, 2].bar_label(axes[1, 2].containers[0])
axes[1, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[2, 1], x=x[6], y=y[6])
axes[2, 1].set_title("Number of Seventh Categories used in Articles of Shahid Beheshti University")
axes[2, 1].bar_label(axes[2, 1].containers[0])
axes[2, 1].tick_params(labelrotation=90)

fig.delaxes(axes[2, 0])
fig.delaxes(axes[2, 2])
```

`trg.right_layout()
plt.show()`



```
In [ ]: fig, axes = plt.subplots(3, 3, figsize=(20, 20))
x = [
    list(df[df['University'] == "Ferdowsi University of Mashhad"]['Category_1'].value),
    list(df[df['University'] == "Ferdowsi University of Mashhad"]['Category_2'].value),
    list(df[df['University'] == "Ferdowsi University of Mashhad"]['Category_3'].value),
    list(df[df['University'] == "Ferdowsi University of Mashhad"]['Category_4'].value),
    list(df[df['University'] == "Ferdowsi University of Mashhad"]['Category_5'].value),
    list(df[df['University'] == "Ferdowsi University of Mashhad"]['Category_6'].value),
    list(df[df['University'] == "Ferdowsi University of Mashhad"]['Category_7'].value)
]
y = [
    list(df[df['University'] == "Ferdowsi University of Mashhad"]['Category_1'].value),
    list(df[df['University'] == "Ferdowsi University of Mashhad"]['Category_2'].value),
    list(df[df['University'] == "Ferdowsi University of Mashhad"]['Category_3'].value),
    list(df[df['University'] == "Ferdowsi University of Mashhad"]['Category_4'].value),
    list(df[df['University'] == "Ferdowsi University of Mashhad"]['Category_5'].value),
    list(df[df['University'] == "Ferdowsi University of Mashhad"]['Category_6'].value),
    list(df[df['University'] == "Ferdowsi University of Mashhad"]['Category_7'].value)
]
sns.barplot(ax=axes[0, 0], x=x[0], y=y[0])
axes[0, 0].set_title("Number of First Categories used in Articles of Ferdowsi University")
axes[0, 0].bar_label(axes[0, 0].containers[0], rotation=90)
axes[0, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 1], x=x[1], y=y[1])
axes[0, 1].set_title("Number of Second Categories used in Articles of Ferdowsi University")
axes[0, 1].bar_label(axes[0, 1].containers[0], rotation=90)
axes[0, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 2], x=x[2], y=y[2])
axes[0, 2].set_title("Number of Third Categories used in Articles of Ferdowsi University")
axes[0, 2].bar_label(axes[0, 2].containers[0], rotation=90)
axes[0, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 0], x=x[3], y=y[3])
axes[1, 0].set_title("Number of Fourth Categories used in Articles of Ferdowsi University")
axes[1, 0].bar_label(axes[1, 0].containers[0])
axes[1, 0].tick_params(labelrotation=90)

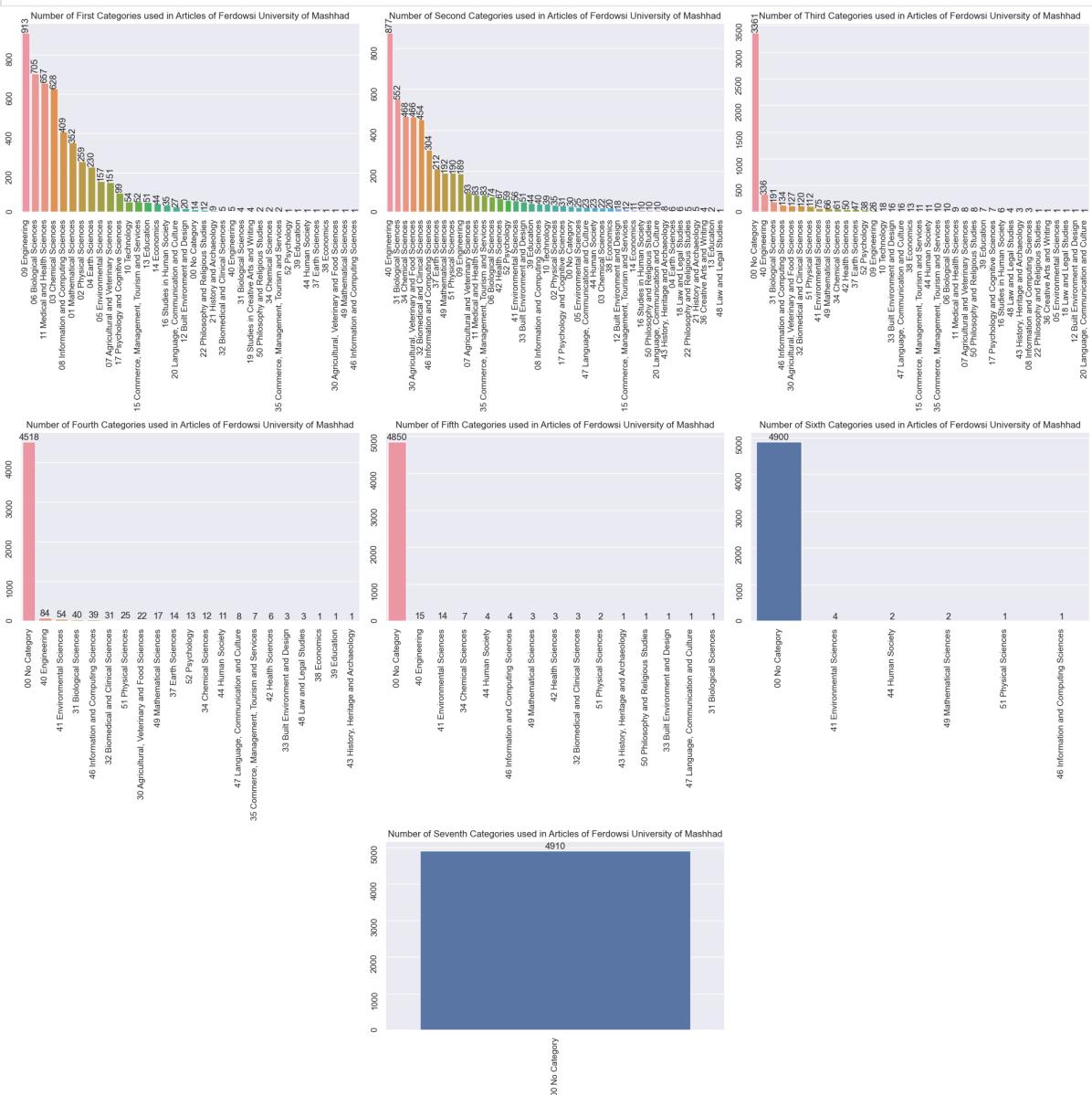
sns.barplot(ax=axes[1, 1], x=x[4], y=y[4])
axes[1, 1].set_title("Number of Fifth Categories used in Articles of Ferdowsi University")
axes[1, 1].bar_label(axes[1, 1].containers[0])
axes[1, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 2], x=x[5], y=y[5])
axes[1, 2].set_title("Number of Sixth Categories used in Articles of Ferdowsi University")
axes[1, 2].bar_label(axes[1, 2].containers[0])
axes[1, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[2, 1], x=x[6], y=y[6])
axes[2, 1].set_title("Number of Seventh Categories used in Articles of Ferdowsi University")
axes[2, 1].bar_label(axes[2, 1].containers[0])
axes[2, 1].tick_params(labelrotation=90)

fig.delaxes(axes[2, 0])
fig.delaxes(axes[2, 2])
```

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



University of Guilan top 7 Categories

- This university also has more evenly distributed articles among categories.
- Still the most published articles from this university are categorized into **Engineering**.

```
In [ ]: fig, axes = plt.subplots(3, 3, figsize=(20, 20))
x = [
    list(df[df['University'] == "University of Guilan"]['Category_1'].value_counts()),
    list(df[df['University'] == "University of Guilan"]['Category_2'].value_counts()),
    list(df[df['University'] == "University of Guilan"]['Category_3'].value_counts()),
    list(df[df['University'] == "University of Guilan"]['Category_4'].value_counts()),
    list(df[df['University'] == "University of Guilan"]['Category_5'].value_counts()),
    list(df[df['University'] == "University of Guilan"]['Category_6'].value_counts()),
    list(df[df['University'] == "University of Guilan"]['Category_7'].value_counts())
]
y = [
    list(df[df['University'] == "University of Guilan"]['Category_1'].value_counts()),
    list(df[df['University'] == "University of Guilan"]['Category_2'].value_counts()),
    list(df[df['University'] == "University of Guilan"]['Category_3'].value_counts()),
    list(df[df['University'] == "University of Guilan"]['Category_4'].value_counts()),
    list(df[df['University'] == "University of Guilan"]['Category_5'].value_counts()),
    list(df[df['University'] == "University of Guilan"]['Category_6'].value_counts()),
    list(df[df['University'] == "University of Guilan"]['Category_7'].value_counts())
]
sns.barplot(ax=axes[0, 0], x=x[0], y=y[0])
axes[0, 0].set_title("Number of First Categories used in Articles of University of
axes[0, 0].bar_label(axes[0, 0].containers[0], rotation=90)
axes[0, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 1], x=x[1], y=y[1])
axes[0, 1].set_title("Number of Second Categories used in Articles of University of
axes[0, 1].bar_label(axes[0, 1].containers[0], rotation=90)
axes[0, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 2], x=x[2], y=y[2])
axes[0, 2].set_title("Number of Third Categories used in Articles of University of
axes[0, 2].bar_label(axes[0, 2].containers[0], rotation=90)
axes[0, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 0], x=x[3], y=y[3])
axes[1, 0].set_title("Number of Fourth Categories used in Articles of University of
axes[1, 0].bar_label(axes[1, 0].containers[0])
axes[1, 0].tick_params(labelrotation=90)

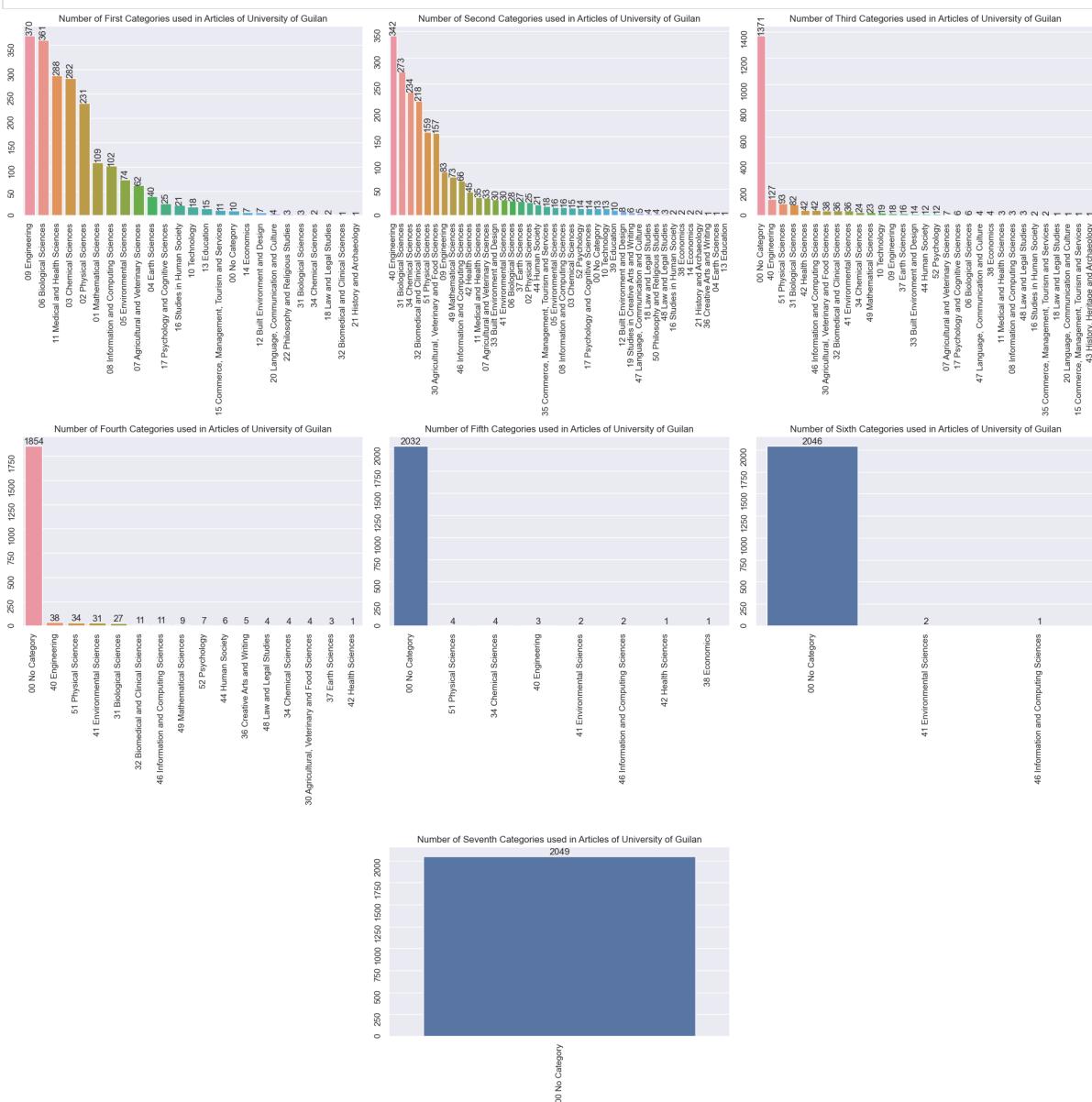
sns.barplot(ax=axes[1, 1], x=x[4], y=y[4])
axes[1, 1].set_title("Number of Fifth Categories used in Articles of University of
axes[1, 1].bar_label(axes[1, 1].containers[0])
axes[1, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 2], x=x[5], y=y[5])
axes[1, 2].set_title("Number of Sixth Categories used in Articles of University of
axes[1, 2].bar_label(axes[1, 2].containers[0])
axes[1, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[2, 1], x=x[6], y=y[6])
axes[2, 1].set_title("Number of Seventh Categories used in Articles of University o
axes[2, 1].bar_label(axes[2, 1].containers[0])
axes[2, 1].tick_params(labelrotation=90)

fig.delaxes(axes[2, 0])
fig.delaxes(axes[2, 2])
```

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



Imam Hossein University top 7 Categories

- This university has the lowest amount of published articles in contrast of other universities.
- Unlike most of the other universities, the most published articles from this university are categorized into **Chemical Sciences**.

```
In [ ]: fig, axes = plt.subplots(3, 3, figsize=(20, 20))
x = [
    list(df[df['University'] == "Imam Hossein University"]['Category_1'].value_coun
    list(df[df['University'] == "Imam Hossein University"]['Category_2'].value_coun
    list(df[df['University'] == "Imam Hossein University"]['Category_3'].value_coun
    list(df[df['University'] == "Imam Hossein University"]['Category_4'].value_coun
    list(df[df['University'] == "Imam Hossein University"]['Category_5'].value_coun
    list(df[df['University'] == "Imam Hossein University"]['Category_6'].value_coun
    list(df[df['University'] == "Imam Hossein University"]['Category_7'].value_coun

]
y = [
    list(df[df['University'] == "Imam Hossein University"]['Category_1'].value_coun
    list(df[df['University'] == "Imam Hossein University"]['Category_2'].value_coun
    list(df[df['University'] == "Imam Hossein University"]['Category_3'].value_coun
    list(df[df['University'] == "Imam Hossein University"]['Category_4'].value_coun
    list(df[df['University'] == "Imam Hossein University"]['Category_5'].value_coun
    list(df[df['University'] == "Imam Hossein University"]['Category_6'].value_coun
    list(df[df['University'] == "Imam Hossein University"]['Category_7'].value_coun

]
sns.barplot(ax=axes[0, 0], x=x[0], y=y[0])
axes[0, 0].set_title("Number of First Categories used in Articles of Imam Hossein U
axes[0, 0].bar_label(axes[0, 0].containers[0], rotation=90)
axes[0, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 1], x=x[1], y=y[1])
axes[0, 1].set_title("Number of Second Categories used in Articles of Imam Hossein U
axes[0, 1].bar_label(axes[0, 1].containers[0], rotation=90)
axes[0, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 2], x=x[2], y=y[2])
axes[0, 2].set_title("Number of Third Categories used in Articles of Imam Hossein U
axes[0, 2].bar_label(axes[0, 2].containers[0], rotation=90)
axes[0, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 0], x=x[3], y=y[3])
axes[1, 0].set_title("Number of Fourth Categories used in Articles of Imam Hossein U
axes[1, 0].bar_label(axes[1, 0].containers[0])
axes[1, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 1], x=x[4], y=y[4])
axes[1, 1].set_title("Number of Fifth Categories used in Articles of Imam Hossein U
axes[1, 1].bar_label(axes[1, 1].containers[0])
axes[1, 1].tick_params(labelrotation=90)

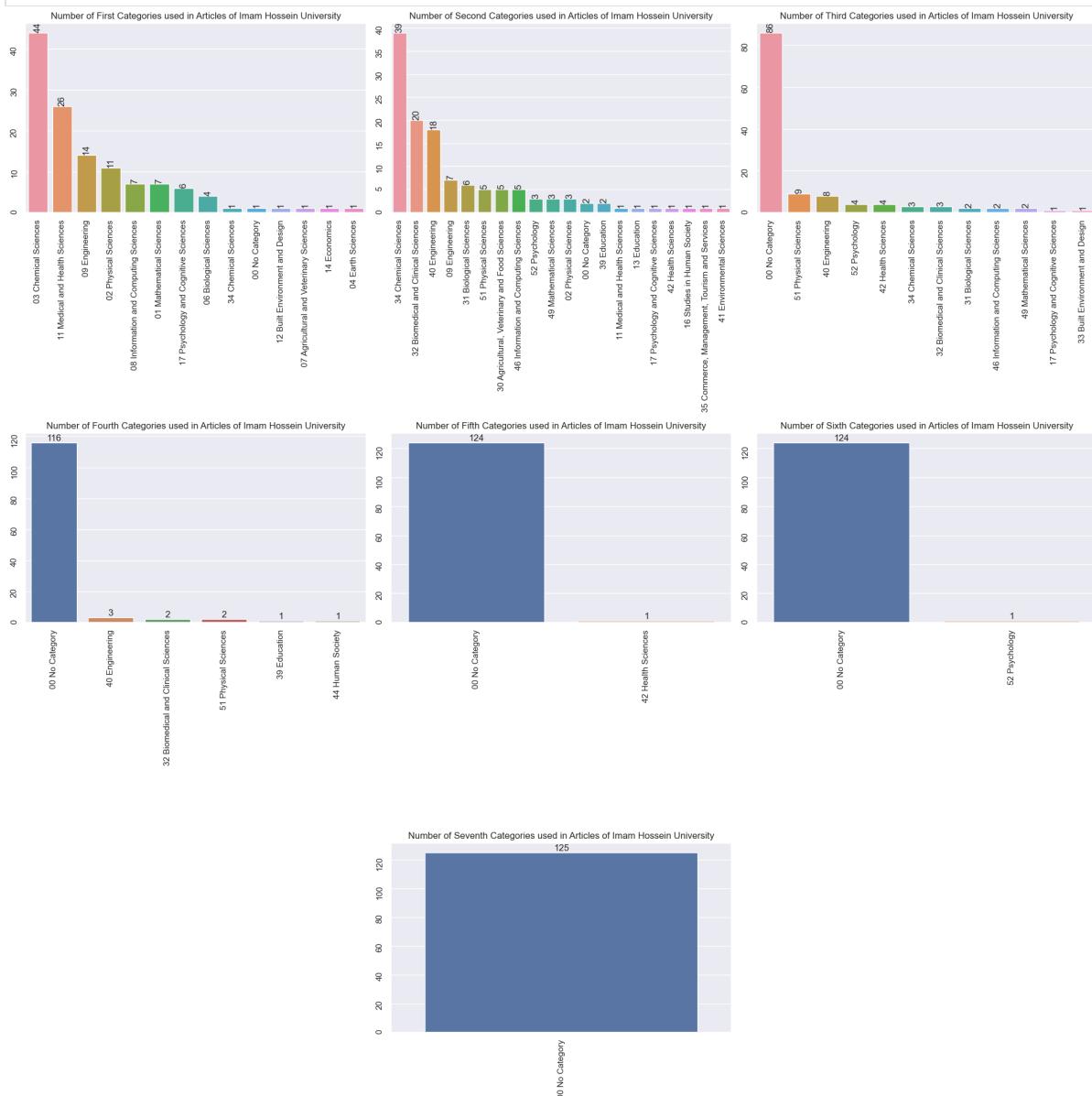
sns.barplot(ax=axes[1, 2], x=x[5], y=y[5])
axes[1, 2].set_title("Number of Sixth Categories used in Articles of Imam Hossein U
axes[1, 2].bar_label(axes[1, 2].containers[0])
axes[1, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[2, 1], x=x[6], y=y[6])
axes[2, 1].set_title("Number of Seventh Categories used in Articles of Imam Hossein U
axes[2, 1].bar_label(axes[2, 1].containers[0])
axes[2, 1].tick_params(labelrotation=90)

fig.delaxes(axes[2, 0])
fig.delaxes(axes[2, 2])


```

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



University of Isfahan top 7 Categories

- The most of articles published by this university is categorized into **Medical and Health Sciences** as their first major category.

```
In [ ]: fig, axes = plt.subplots(3, 3, figsize=(20, 20))
x = [
    list(df[df['University'] == "University of Isfahan"]['Category_1'].value_counts),
    list(df[df['University'] == "University of Isfahan"]['Category_2'].value_counts),
    list(df[df['University'] == "University of Isfahan"]['Category_3'].value_counts),
    list(df[df['University'] == "University of Isfahan"]['Category_4'].value_counts),
    list(df[df['University'] == "University of Isfahan"]['Category_5'].value_counts),
    list(df[df['University'] == "University of Isfahan"]['Category_6'].value_counts),
    list(df[df['University'] == "University of Isfahan"]['Category_7'].value_counts)
]
y = [
    list(df[df['University'] == "University of Isfahan"]['Category_1'].value_counts),
    list(df[df['University'] == "University of Isfahan"]['Category_2'].value_counts),
    list(df[df['University'] == "University of Isfahan"]['Category_3'].value_counts),
    list(df[df['University'] == "University of Isfahan"]['Category_4'].value_counts),
    list(df[df['University'] == "University of Isfahan"]['Category_5'].value_counts),
    list(df[df['University'] == "University of Isfahan"]['Category_6'].value_counts),
    list(df[df['University'] == "University of Isfahan"]['Category_7'].value_counts)
]
sns.barplot(ax=axes[0, 0], x=x[0], y=y[0])
axes[0, 0].set_title("Number of First Categories used in Articles of University of Isfahan")
axes[0, 0].bar_label(axes[0, 0].containers[0], rotation=90)
axes[0, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 1], x=x[1], y=y[1])
axes[0, 1].set_title("Number of Second Categories used in Articles of University of Isfahan")
axes[0, 1].bar_label(axes[0, 1].containers[0], rotation=90)
axes[0, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 2], x=x[2], y=y[2])
axes[0, 2].set_title("Number of Third Categories used in Articles of University of Isfahan")
axes[0, 2].bar_label(axes[0, 2].containers[0], rotation=90)
axes[0, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 0], x=x[3], y=y[3])
axes[1, 0].set_title("Number of Fourth Categories used in Articles of University of Isfahan")
axes[1, 0].bar_label(axes[1, 0].containers[0])
axes[1, 0].tick_params(labelrotation=90)

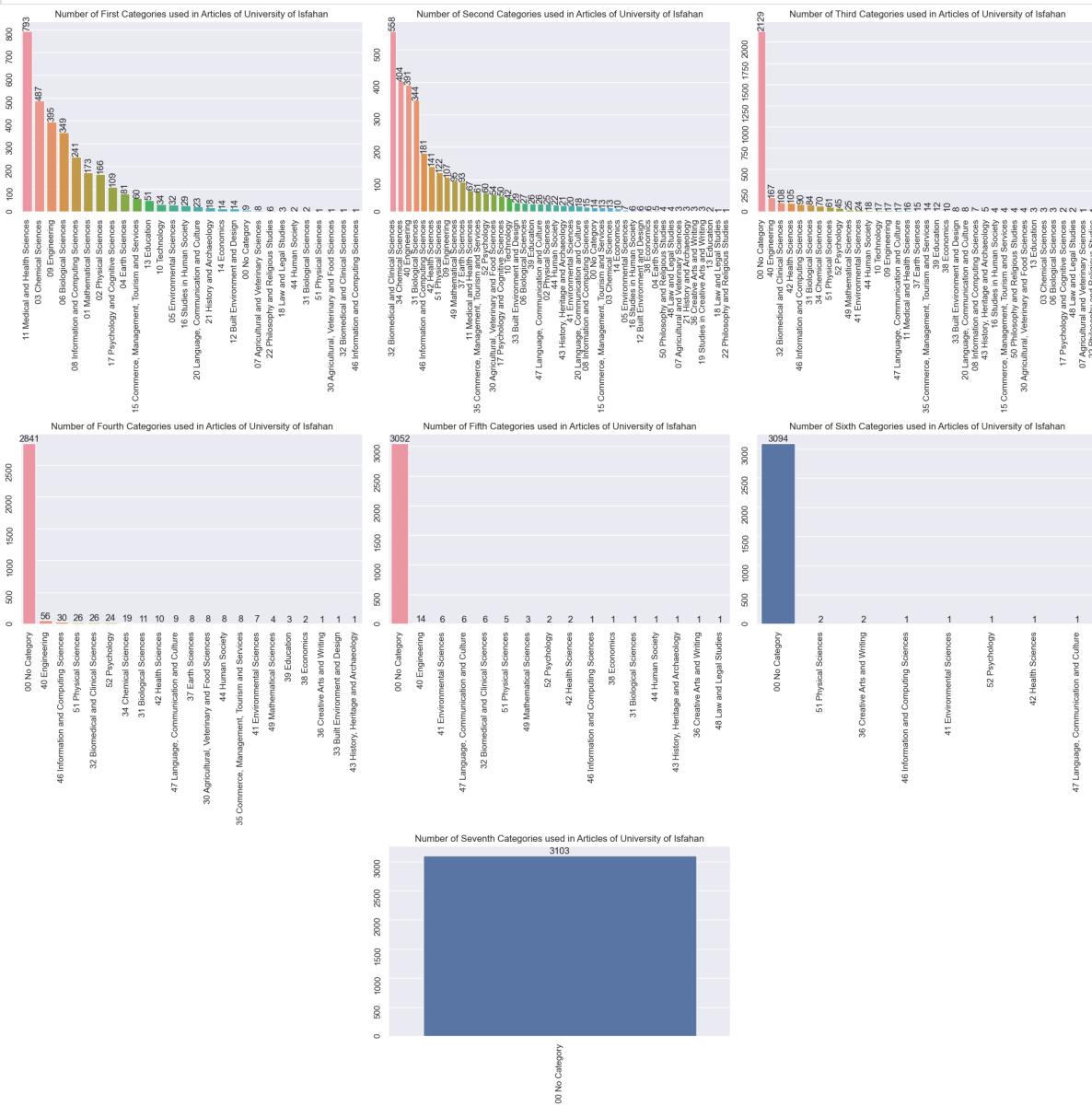
sns.barplot(ax=axes[1, 1], x=x[4], y=y[4])
axes[1, 1].set_title("Number of Fifth Categories used in Articles of University of Isfahan")
axes[1, 1].bar_label(axes[1, 1].containers[0])
axes[1, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 2], x=x[5], y=y[5])
axes[1, 2].set_title("Number of Sixth Categories used in Articles of University of Isfahan")
axes[1, 2].bar_label(axes[1, 2].containers[0])
axes[1, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[2, 1], x=x[6], y=y[6])
axes[2, 1].set_title("Number of Seventh Categories used in Articles of University of Isfahan")
axes[2, 1].bar_label(axes[2, 1].containers[0])
axes[2, 1].tick_params(labelrotation=90)

fig.delaxes(axes[2, 0])
fig.delaxes(axes[2, 2])
```

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



Isfahan University of Technology top 7 Categories

- As it was predictable, since this university is more focused on the technical sciences, the most published articles of this university, categorized into **Engineering** as their first major category.
- the **Engineering** category as first or second major one, has a significant distance from other categories.

```
In [ ]: fig, axes = plt.subplots(3, 3, figsize=(20, 20))
x = [
    list(df[df['University'] == "Isfahan University of Technology"]['Category_1'].v
    list(df[df['University'] == "Isfahan University of Technology"]['Category_2'].v
    list(df[df['University'] == "Isfahan University of Technology"]['Category_3'].v
    list(df[df['University'] == "Isfahan University of Technology"]['Category_4'].v
    list(df[df['University'] == "Isfahan University of Technology"]['Category_5'].v
    list(df[df['University'] == "Isfahan University of Technology"]['Category_6'].v
    list(df[df['University'] == "Isfahan University of Technology"]['Category_7'].v
]

]
y = [
    list(df[df['University'] == "Isfahan University of Technology"]['Category_1'].v
    list(df[df['University'] == "Isfahan University of Technology"]['Category_2'].v
    list(df[df['University'] == "Isfahan University of Technology"]['Category_3'].v
    list(df[df['University'] == "Isfahan University of Technology"]['Category_4'].v
    list(df[df['University'] == "Isfahan University of Technology"]['Category_5'].v
    list(df[df['University'] == "Isfahan University of Technology"]['Category_6'].v
    list(df[df['University'] == "Isfahan University of Technology"]['Category_7'].v
]

]
sns.barplot(ax=axes[0, 0], x=x[0], y=y[0])
axes[0, 0].set_title("Number of First Categories used in Articles of Isfahan Univer
axes[0, 0].bar_label(axes[0, 0].containers[0], rotation=90)
axes[0, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 1], x=x[1], y=y[1])
axes[0, 1].set_title("Number of Second Categories used in Articles of Isfahan Unive
axes[0, 1].bar_label(axes[0, 1].containers[0], rotation=90)
axes[0, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 2], x=x[2], y=y[2])
axes[0, 2].set_title("Number of Third Categories used in Articles of Isfahan Univer
axes[0, 2].bar_label(axes[0, 2].containers[0], rotation=90)
axes[0, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 0], x=x[3], y=y[3])
axes[1, 0].set_title("Number of Fourth Categories used in Articles of Isfahan Unive
axes[1, 0].bar_label(axes[1, 0].containers[0])
axes[1, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 1], x=x[4], y=y[4])
axes[1, 1].set_title("Number of Fifth Categories used in Articles of Isfahan Univer
axes[1, 1].bar_label(axes[1, 1].containers[0])
axes[1, 1].tick_params(labelrotation=90)

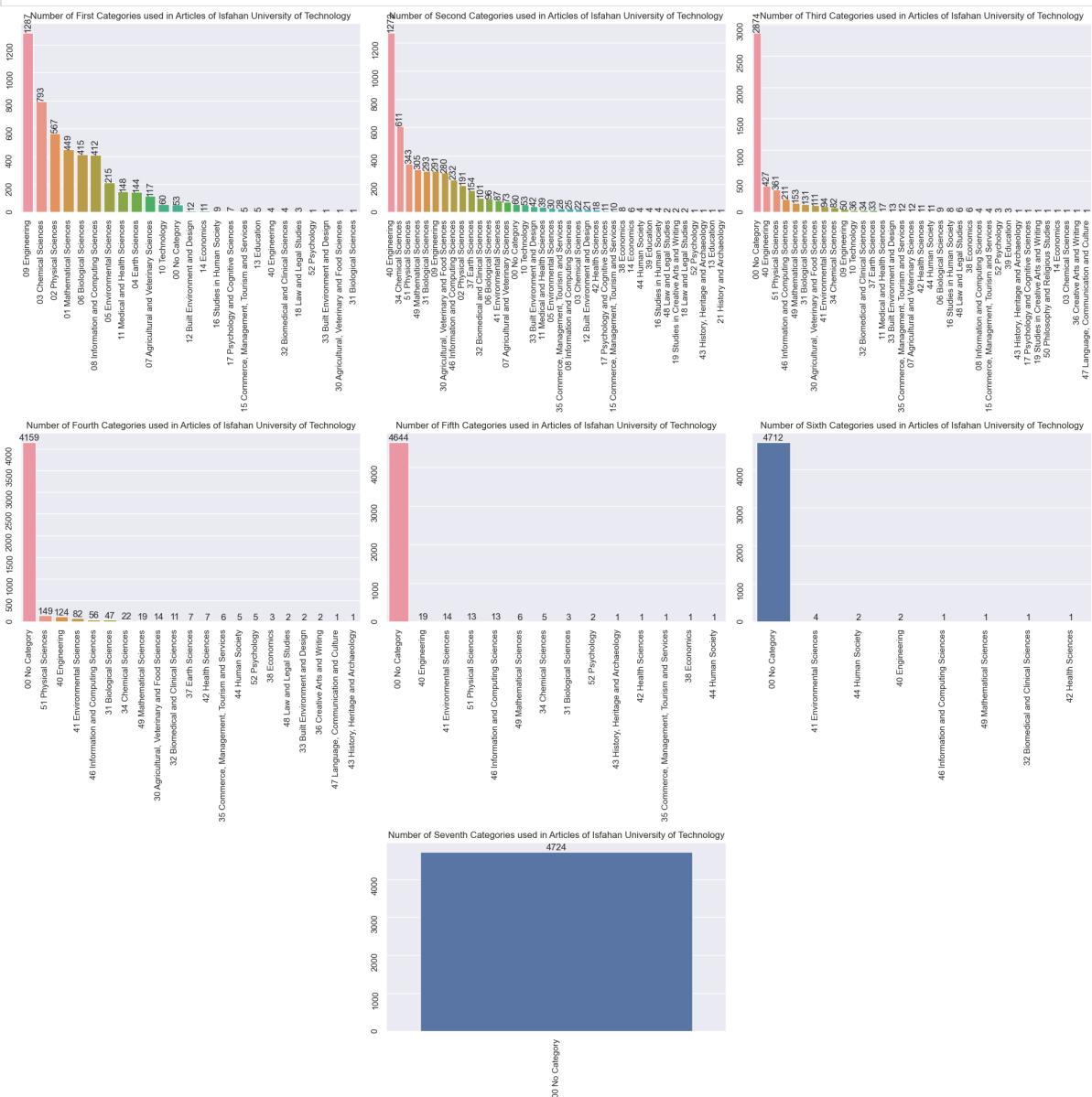
sns.barplot(ax=axes[1, 2], x=x[5], y=y[5])
axes[1, 2].set_title("Number of Sixth Categories used in Articles of Isfahan Univer
axes[1, 2].bar_label(axes[1, 2].containers[0])
axes[1, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[2, 1], x=x[6], y=y[6])
axes[2, 1].set_title("Number of Seventh Categories used in Articles of Isfahan Univ
axes[2, 1].bar_label(axes[2, 1].containers[0])
axes[2, 1].tick_params(labelrotation=90)

fig.delaxes(axes[2, 0])
fig.delaxes(axes[2, 2])


```

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



K. N. Toosi University of Technology top 7 Categories

- As it was predictable, since this university is more focused on the technical sciences, the most published articles of this university, categorized into **Engineering** as their first major category.
- the **Engineering** category as first or second major one, has a significant distance from other categories.

```
In [ ]: fig, axes = plt.subplots(3, 3, figsize=(20, 20))
x = [
    list(df[df['University'] == "K. N. Toosi University of Technology"])[['Category_1',
    list(df[df['University'] == "K. N. Toosi University of Technology"])[['Category_2',
    list(df[df['University'] == "K. N. Toosi University of Technology"])[['Category_3',
    list(df[df['University'] == "K. N. Toosi University of Technology"])[['Category_4',
    list(df[df['University'] == "K. N. Toosi University of Technology"])[['Category_5',
    list(df[df['University'] == "K. N. Toosi University of Technology"])[['Category_6',
    list(df[df['University'] == "K. N. Toosi University of Technology"])[['Category_7',
    ],
    y = [
        list(df[df['University'] == "K. N. Toosi University of Technology"])[['Category_1',
        list(df[df['University'] == "K. N. Toosi University of Technology"])[['Category_2',
        list(df[df['University'] == "K. N. Toosi University of Technology"])[['Category_3',
        list(df[df['University'] == "K. N. Toosi University of Technology"])[['Category_4',
        list(df[df['University'] == "K. N. Toosi University of Technology"])[['Category_5',
        list(df[df['University'] == "K. N. Toosi University of Technology"])[['Category_6',
        list(df[df['University'] == "K. N. Toosi University of Technology"])[['Category_7',
        ],
        sns.barplot(ax=axes[0, 0], x=x[0], y=y[0])
        axes[0, 0].set_title("Number of First Categories used in Articles of K. N. Toosi Un")
        axes[0, 0].bar_label(axes[0, 0].containers[0], rotation=90)
        axes[0, 0].tick_params(labelrotation=90)

        sns.barplot(ax=axes[0, 1], x=x[1], y=y[1])
        axes[0, 1].set_title("Number of Second Categories used in Articles of K. N. Toosi U")
        axes[0, 1].bar_label(axes[0, 1].containers[0], rotation=90)
        axes[0, 1].tick_params(labelrotation=90)

        sns.barplot(ax=axes[0, 2], x=x[2], y=y[2])
        axes[0, 2].set_title("Number of Third Categories used in Articles of K. N. Toosi Un")
        axes[0, 2].bar_label(axes[0, 2].containers[0], rotation=90)
        axes[0, 2].tick_params(labelrotation=90)

        sns.barplot(ax=axes[1, 0], x=x[3], y=y[3])
        axes[1, 0].set_title("Number of Fourth Categories used in Articles of K. N. Toosi Un")
        axes[1, 0].bar_label(axes[1, 0].containers[0])
        axes[1, 0].tick_params(labelrotation=90)

        sns.barplot(ax=axes[1, 1], x=x[4], y=y[4])
        axes[1, 1].set_title("Number of Fifth Categories used in Articles of K. N. Toosi Un")
        axes[1, 1].bar_label(axes[1, 1].containers[0])
        axes[1, 1].tick_params(labelrotation=90)

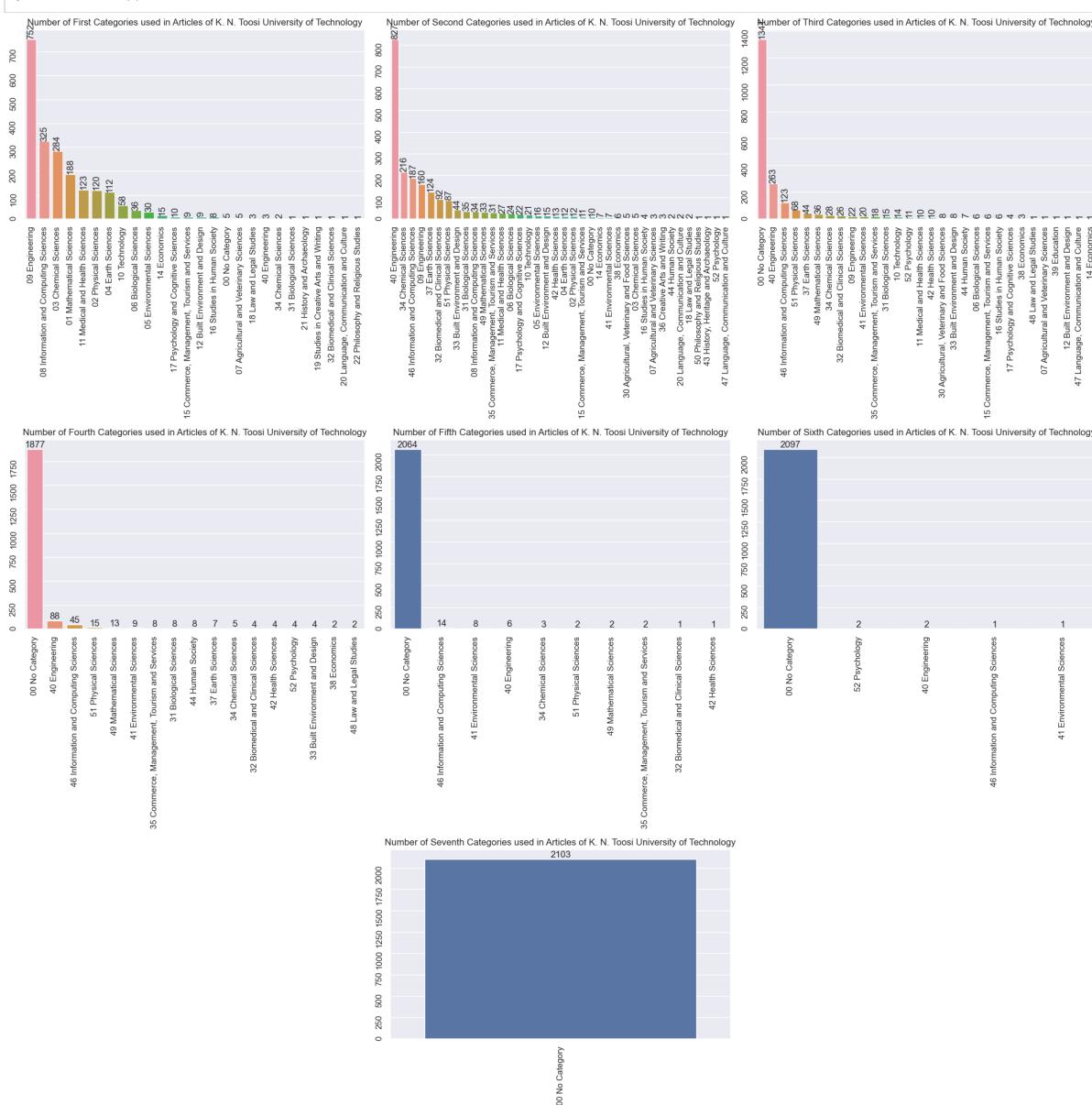
        sns.barplot(ax=axes[1, 2], x=x[5], y=y[5])
        axes[1, 2].set_title("Number of Sixth Categories used in Articles of K. N. Toosi Un")
        axes[1, 2].bar_label(axes[1, 2].containers[0])
        axes[1, 2].tick_params(labelrotation=90)

        sns.barplot(ax=axes[2, 1], x=x[6], y=y[6])
        axes[2, 1].set_title("Number of Seventh Categories used in Articles of K. N. Toosi Un")
        axes[2, 1].bar_label(axes[2, 1].containers[0])
        axes[2, 1].tick_params(labelrotation=90)

fig.delaxes(axes[2, 0])
fig.delaxes(axes[2, 2])

fig.tight_layout()
```

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



Sharif University of Technology top 7 Categories

- Although labeled as the second best university in Iran, its research output is not in the second place, it's in third place after *Tarbiat Modares University*.
- As it was predictable, since this university is more focused on the technical sciences, the most published articles of this university, categorized into **Engineering** as their first major category.
- the **Engineering** category as first or second major one, has a significant distance from other categories.

```
In [ ]: fig, axes = plt.subplots(3, 3, figsize=(20, 20))
x = [
    list(df[df['University'] == "Sharif University of Technology"]['Category_1'].value_counts()),
    list(df[df['University'] == "Sharif University of Technology"]['Category_2'].value_counts()),
    list(df[df['University'] == "Sharif University of Technology"]['Category_3'].value_counts()),
    list(df[df['University'] == "Sharif University of Technology"]['Category_4'].value_counts()),
    list(df[df['University'] == "Sharif University of Technology"]['Category_5'].value_counts()),
    list(df[df['University'] == "Sharif University of Technology"]['Category_6'].value_counts()),
    list(df[df['University'] == "Sharif University of Technology"]['Category_7'].value_counts())
]
y = [
    list(df[df['University'] == "Sharif University of Technology"]['Category_1'].value_counts()),
    list(df[df['University'] == "Sharif University of Technology"]['Category_2'].value_counts()),
    list(df[df['University'] == "Sharif University of Technology"]['Category_3'].value_counts()),
    list(df[df['University'] == "Sharif University of Technology"]['Category_4'].value_counts()),
    list(df[df['University'] == "Sharif University of Technology"]['Category_5'].value_counts()),
    list(df[df['University'] == "Sharif University of Technology"]['Category_6'].value_counts()),
    list(df[df['University'] == "Sharif University of Technology"]['Category_7'].value_counts())
]
sns.barplot(ax=axes[0, 0], x=x[0], y=y[0])
axes[0, 0].set_title("Number of First Categories used in Articles of Sharif University")
axes[0, 0].bar_label(axes[0, 0].containers[0], rotation=90)
axes[0, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 1], x=x[1], y=y[1])
axes[0, 1].set_title("Number of Second Categories used in Articles of Sharif University")
axes[0, 1].bar_label(axes[0, 1].containers[0], rotation=90)
axes[0, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 2], x=x[2], y=y[2])
axes[0, 2].set_title("Number of Third Categories used in Articles of Sharif University")
axes[0, 2].bar_label(axes[0, 2].containers[0], rotation=90)
axes[0, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 0], x=x[3], y=y[3])
axes[1, 0].set_title("Number of Fourth Categories used in Articles of Sharif University")
axes[1, 0].bar_label(axes[1, 0].containers[0])
axes[1, 0].tick_params(labelrotation=90)

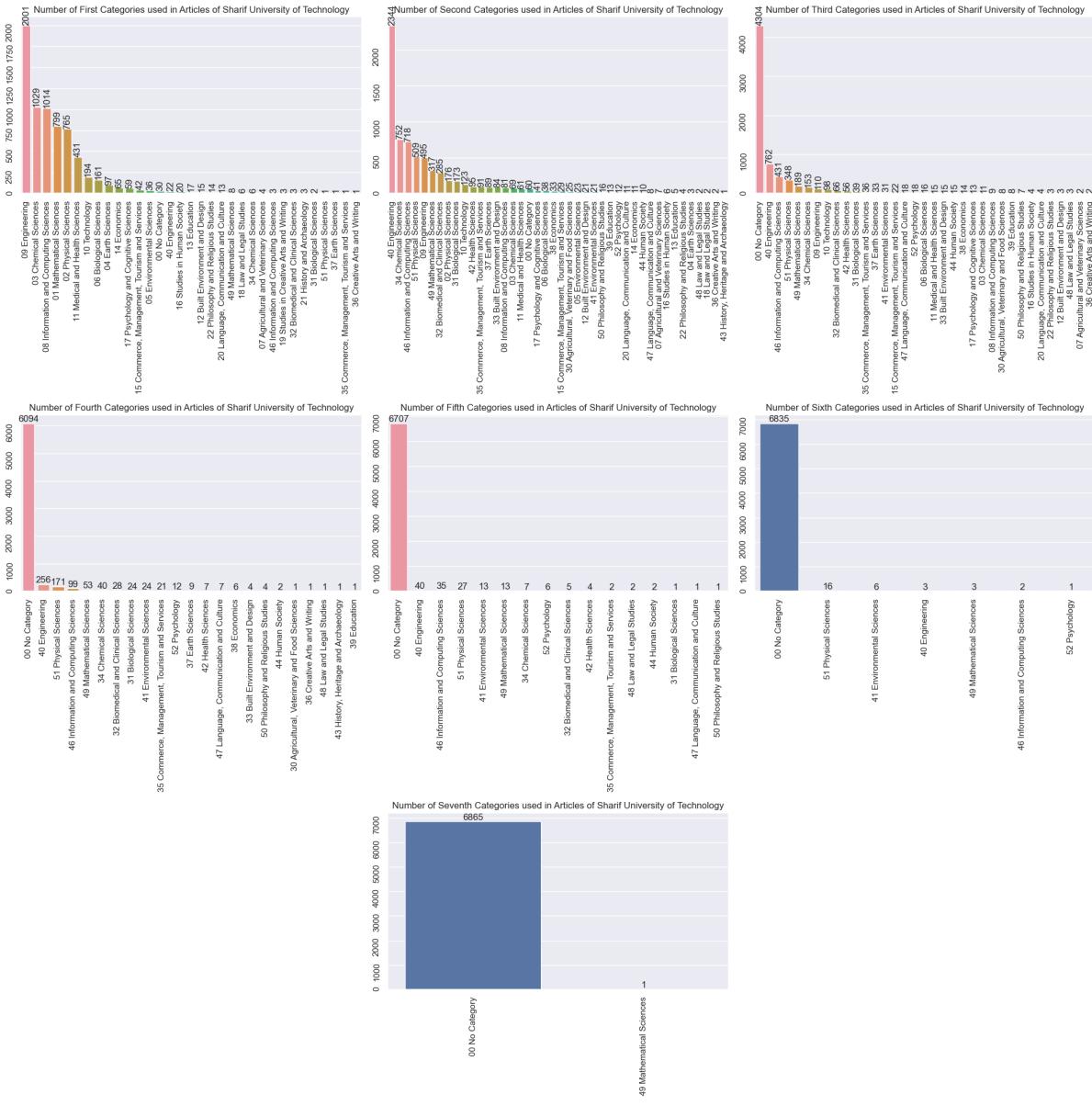
sns.barplot(ax=axes[1, 1], x=x[4], y=y[4])
axes[1, 1].set_title("Number of Fifth Categories used in Articles of Sharif University")
axes[1, 1].bar_label(axes[1, 1].containers[0])
axes[1, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 2], x=x[5], y=y[5])
axes[1, 2].set_title("Number of Sixth Categories used in Articles of Sharif University")
axes[1, 2].bar_label(axes[1, 2].containers[0])
axes[1, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[2, 1], x=x[6], y=y[6])
axes[2, 1].set_title("Number of Seventh Categories used in Articles of Sharif University")
axes[2, 1].bar_label(axes[2, 1].containers[0])
axes[2, 1].tick_params(labelrotation=90)

fig.delaxes(axes[2, 0])
fig.delaxes(axes[2, 2])
```

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



University of Shiraz Top 7 Categories

- **Engineering** is the top major category for the published articles of this university.
- **Biological Sciences, Medical and Health Sciences** and **Chemical Sciences** are in the second place with almost similar amount.

```
In [ ]: fig, axes = plt.subplots(3, 3, figsize=(20, 20))
x = [
    list(df[df['University'] == "University of Shiraz"]['Category_1'].value_counts()),
    list(df[df['University'] == "University of Shiraz"]['Category_2'].value_counts()),
    list(df[df['University'] == "University of Shiraz"]['Category_3'].value_counts()),
    list(df[df['University'] == "University of Shiraz"]['Category_4'].value_counts()),
    list(df[df['University'] == "University of Shiraz"]['Category_5'].value_counts()),
    list(df[df['University'] == "University of Shiraz"]['Category_6'].value_counts()),
    list(df[df['University'] == "University of Shiraz"]['Category_7'].value_counts())
]
y = [
    list(df[df['University'] == "University of Shiraz"]['Category_1'].value_counts()),
    list(df[df['University'] == "University of Shiraz"]['Category_2'].value_counts()),
    list(df[df['University'] == "University of Shiraz"]['Category_3'].value_counts()),
    list(df[df['University'] == "University of Shiraz"]['Category_4'].value_counts()),
    list(df[df['University'] == "University of Shiraz"]['Category_5'].value_counts()),
    list(df[df['University'] == "University of Shiraz"]['Category_6'].value_counts()),
    list(df[df['University'] == "University of Shiraz"]['Category_7'].value_counts())
]
sns.barplot(ax=axes[0, 0], x=x[0], y=y[0])
axes[0, 0].set_title("Number of First Categories used in Articles of University of")
axes[0, 0].bar_label(axes[0, 0].containers[0], rotation=90)
axes[0, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 1], x=x[1], y=y[1])
axes[0, 1].set_title("Number of Second Categories used in Articles of University of")
axes[0, 1].bar_label(axes[0, 1].containers[0], rotation=90)
axes[0, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 2], x=x[2], y=y[2])
axes[0, 2].set_title("Number of Third Categories used in Articles of University of")
axes[0, 2].bar_label(axes[0, 2].containers[0], rotation=90)
axes[0, 2].tick_params(labelrotation=90)

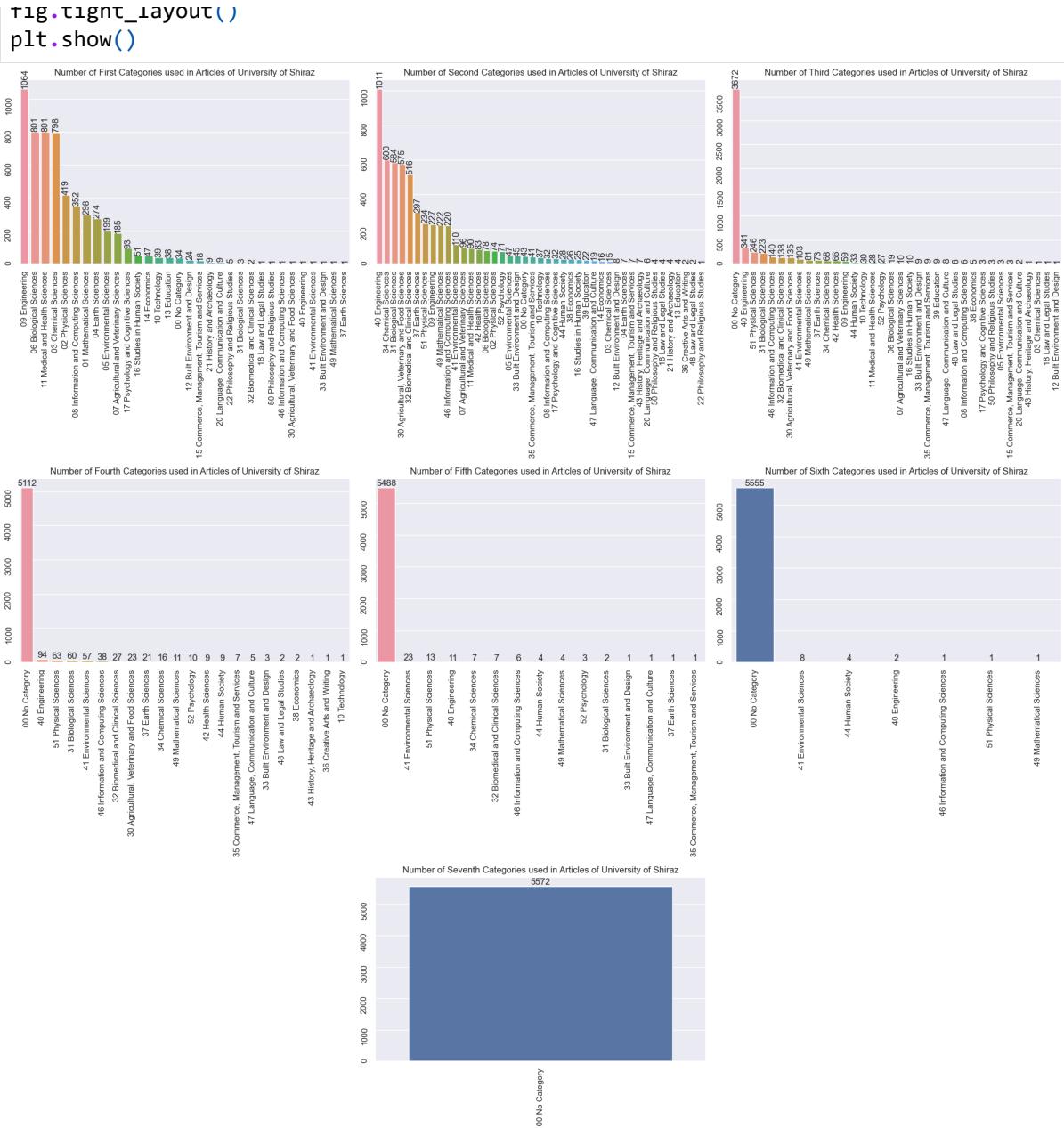
sns.barplot(ax=axes[1, 0], x=x[3], y=y[3])
axes[1, 0].set_title("Number of Fourth Categories used in Articles of University of")
axes[1, 0].bar_label(axes[1, 0].containers[0])
axes[1, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 1], x=x[4], y=y[4])
axes[1, 1].set_title("Number of Fifth Categories used in Articles of University of")
axes[1, 1].bar_label(axes[1, 1].containers[0])
axes[1, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 2], x=x[5], y=y[5])
axes[1, 2].set_title("Number of Sixth Categories used in Articles of University of")
axes[1, 2].bar_label(axes[1, 2].containers[0])
axes[1, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[2, 1], x=x[6], y=y[6])
axes[2, 1].set_title("Number of Seventh Categories used in Articles of University o")
axes[2, 1].bar_label(axes[2, 1].containers[0])
axes[2, 1].tick_params(labelrotation=90)

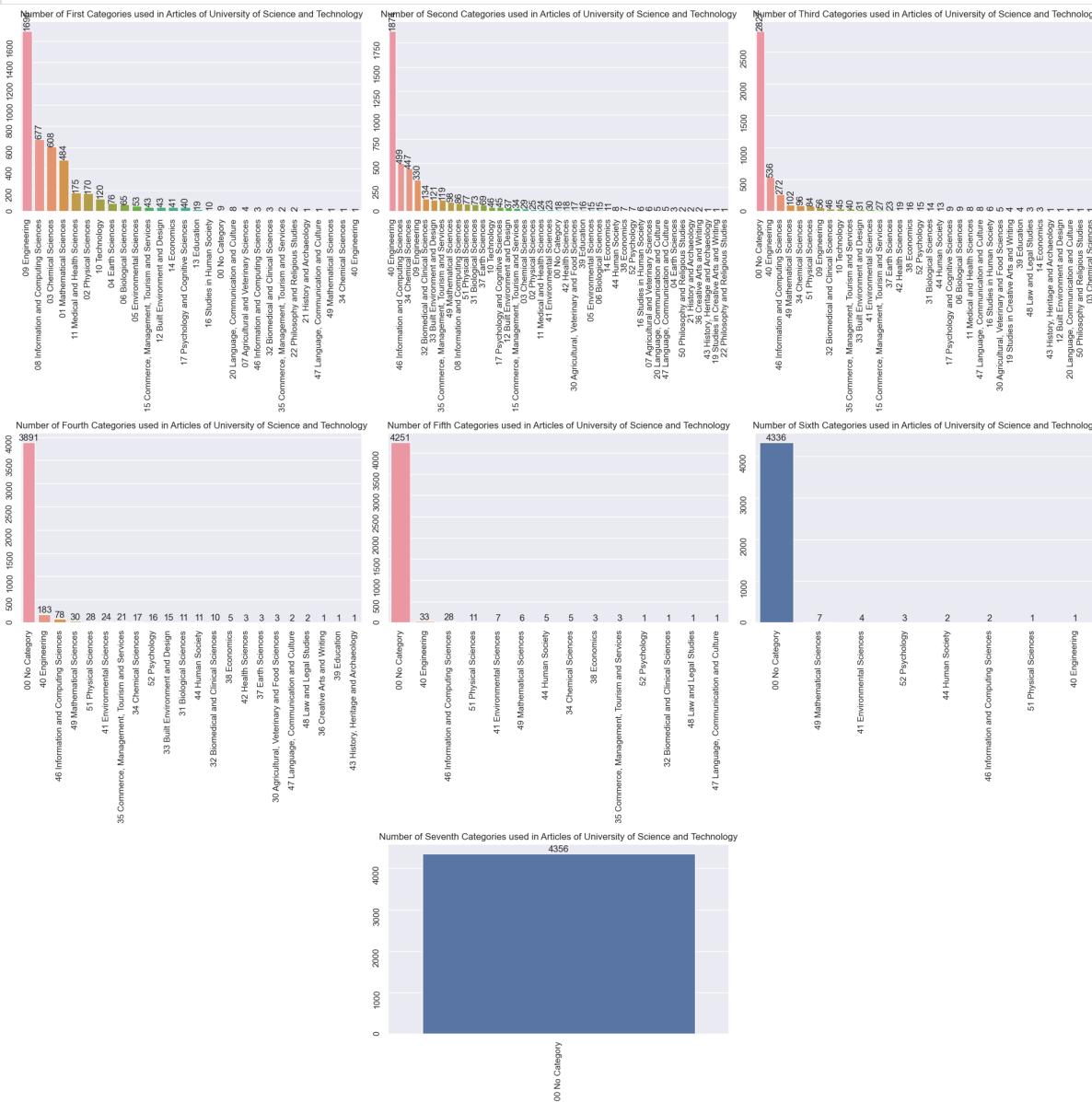
fig.delaxes(axes[2, 0])
fig.delaxes(axes[2, 2])
```



Iran University of Science and Technology Top 7 Categories

- As it was predictable, since this university is more focused on the technical sciences, the most published articles of this university, categorized into **Engineering** as their first major category.
- the **Engineering** category as first or second major one, has a significant distance from other categories.

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



University of Tabriz Top 7 Categories

- **Engineering** is the top major category for the published articles of this university.

```
In [ ]: fig, axes = plt.subplots(3, 3, figsize=(20, 20))
x = [
    list(df[df['University'] == "University of Tabriz"]['Category_1'].value_counts()),
    list(df[df['University'] == "University of Tabriz"]['Category_2'].value_counts()),
    list(df[df['University'] == "University of Tabriz"]['Category_3'].value_counts()),
    list(df[df['University'] == "University of Tabriz"]['Category_4'].value_counts()),
    list(df[df['University'] == "University of Tabriz"]['Category_5'].value_counts()),
    list(df[df['University'] == "University of Tabriz"]['Category_6'].value_counts()),
    list(df[df['University'] == "University of Tabriz"]['Category_7'].value_counts())
]
y = [
    list(df[df['University'] == "University of Tabriz"]['Category_1'].value_counts()),
    list(df[df['University'] == "University of Tabriz"]['Category_2'].value_counts()),
    list(df[df['University'] == "University of Tabriz"]['Category_3'].value_counts()),
    list(df[df['University'] == "University of Tabriz"]['Category_4'].value_counts()),
    list(df[df['University'] == "University of Tabriz"]['Category_5'].value_counts()),
    list(df[df['University'] == "University of Tabriz"]['Category_6'].value_counts()),
    list(df[df['University'] == "University of Tabriz"]['Category_7'].value_counts())
]
sns.barplot(ax=axes[0, 0], x=x[0], y=y[0])
axes[0, 0].set_title("Number of First Categories used in Articles of University of")
axes[0, 0].bar_label(axes[0, 0].containers[0], rotation=90)
axes[0, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 1], x=x[1], y=y[1])
axes[0, 1].set_title("Number of Second Categories used in Articles of University of")
axes[0, 1].bar_label(axes[0, 1].containers[0], rotation=90)
axes[0, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 2], x=x[2], y=y[2])
axes[0, 2].set_title("Number of Third Categories used in Articles of University of")
axes[0, 2].bar_label(axes[0, 2].containers[0], rotation=90)
axes[0, 2].tick_params(labelrotation=90)

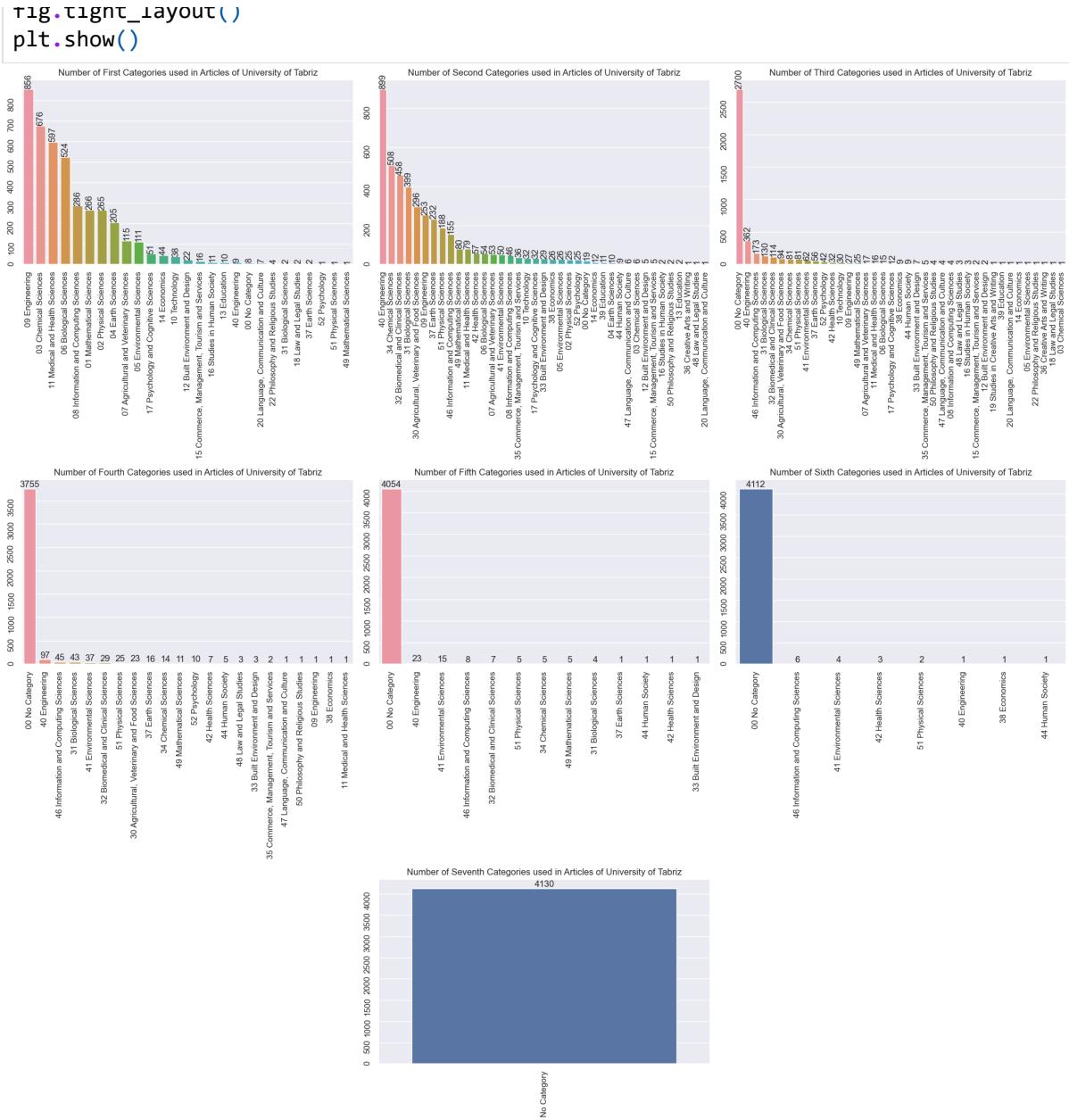
sns.barplot(ax=axes[1, 0], x=x[3], y=y[3])
axes[1, 0].set_title("Number of Fourth Categories used in Articles of University of")
axes[1, 0].bar_label(axes[1, 0].containers[0])
axes[1, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 1], x=x[4], y=y[4])
axes[1, 1].set_title("Number of Fifth Categories used in Articles of University of")
axes[1, 1].bar_label(axes[1, 1].containers[0])
axes[1, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 2], x=x[5], y=y[5])
axes[1, 2].set_title("Number of Sixth Categories used in Articles of University of")
axes[1, 2].bar_label(axes[1, 2].containers[0])
axes[1, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[2, 1], x=x[6], y=y[6])
axes[2, 1].set_title("Number of Seventh Categories used in Articles of University o")
axes[2, 1].bar_label(axes[2, 1].containers[0])
axes[2, 1].tick_params(labelrotation=90)

fig.delaxes(axes[2, 0])
fig.delaxes(axes[2, 2])
```



Tarbiat Modares University Top 7 Categories

- This university has the second most published articles among the investigated universities.
 - Unlike the majority of other universities, **Medical and Health Sciences** is the most used category as the first major category of published articles of this university.

```
In [ ]: fig, axes = plt.subplots(3, 3, figsize=(20, 20))
x = [
    list(df[df['University'] == "Tarbiat Modares University"]['Category_1'].value_c),
    list(df[df['University'] == "Tarbiat Modares University"]['Category_2'].value_c),
    list(df[df['University'] == "Tarbiat Modares University"]['Category_3'].value_c),
    list(df[df['University'] == "Tarbiat Modares University"]['Category_4'].value_c),
    list(df[df['University'] == "Tarbiat Modares University"]['Category_5'].value_c),
    list(df[df['University'] == "Tarbiat Modares University"]['Category_6'].value_c),
    list(df[df['University'] == "Tarbiat Modares University"]['Category_7'].value_c)
]
y = [
    list(df[df['University'] == "Tarbiat Modares University"]['Category_1'].value_c),
    list(df[df['University'] == "Tarbiat Modares University"]['Category_2'].value_c),
    list(df[df['University'] == "Tarbiat Modares University"]['Category_3'].value_c),
    list(df[df['University'] == "Tarbiat Modares University"]['Category_4'].value_c),
    list(df[df['University'] == "Tarbiat Modares University"]['Category_5'].value_c),
    list(df[df['University'] == "Tarbiat Modares University"]['Category_6'].value_c),
    list(df[df['University'] == "Tarbiat Modares University"]['Category_7'].value_c)
]
sns.barplot(ax=axes[0, 0], x=x[0], y=y[0])
axes[0, 0].set_title("Number of First Categories used in Articles of Tarbiat Modare")
axes[0, 0].bar_label(axes[0, 0].containers[0], rotation=90)
axes[0, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 1], x=x[1], y=y[1])
axes[0, 1].set_title("Number of Second Categories used in Articles of Tarbiat Modar")
axes[0, 1].bar_label(axes[0, 1].containers[0], rotation=90)
axes[0, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 2], x=x[2], y=y[2])
axes[0, 2].set_title("Number of Third Categories used in Articles of Tarbiat Modare")
axes[0, 2].bar_label(axes[0, 2].containers[0], rotation=90)
axes[0, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 0], x=x[3], y=y[3])
axes[1, 0].set_title("Number of Fourth Categories used in Articles of Tarbiat Modar")
axes[1, 0].bar_label(axes[1, 0].containers[0])
axes[1, 0].tick_params(labelrotation=90)

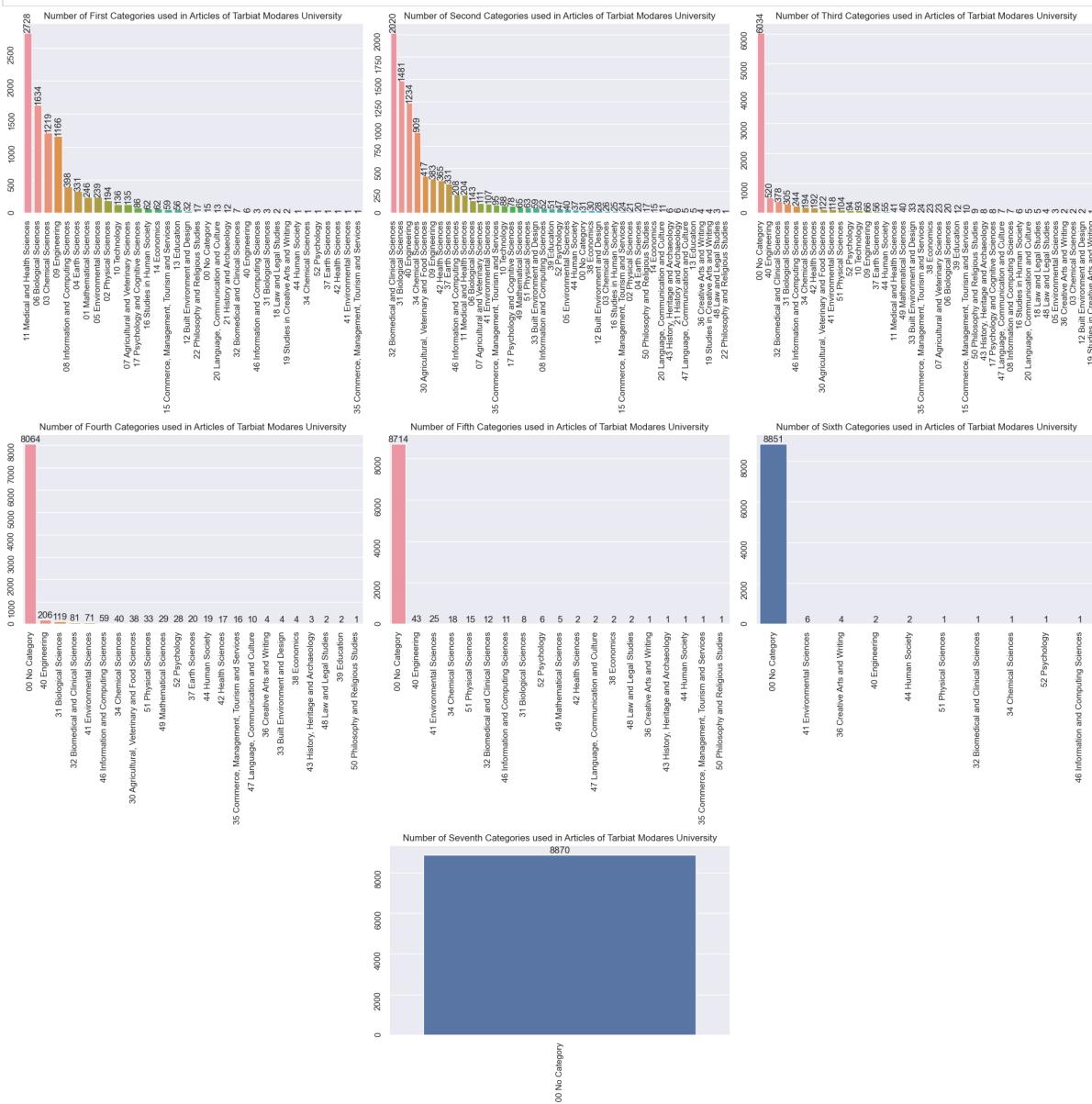
sns.barplot(ax=axes[1, 1], x=x[4], y=y[4])
axes[1, 1].set_title("Number of Fifth Categories used in Articles of Tarbiat Modare")
axes[1, 1].bar_label(axes[1, 1].containers[0])
axes[1, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 2], x=x[5], y=y[5])
axes[1, 2].set_title("Number of Sixth Categories used in Articles of Tarbiat Modare")
axes[1, 2].bar_label(axes[1, 2].containers[0])
axes[1, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[2, 1], x=x[6], y=y[6])
axes[2, 1].set_title("Number of Seventh Categories used in Articles of Tarbiat Moda")
axes[2, 1].bar_label(axes[2, 1].containers[0])
axes[2, 1].tick_params(labelrotation=90)

fig.delaxes(axes[2, 0])
fig.delaxes(axes[2, 2])
```

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



University of Tehran Top 7 Categories

- This university has the most published articles among the universities and ranked as the top university of Iran.
- Although the most used first major of this university published articles is **Engineering**, for each specialized categories such as human sciences and etc, still has the most published articles in contrast of other niche focused universities.

```
In [ ]: fig, axes = plt.subplots(3, 3, figsize=(20, 20))
x = [
    list(df[df['University'] == "University of Tehran"]['Category_1'].value_counts()),
    list(df[df['University'] == "University of Tehran"]['Category_2'].value_counts()),
    list(df[df['University'] == "University of Tehran"]['Category_3'].value_counts()),
    list(df[df['University'] == "University of Tehran"]['Category_4'].value_counts()),
    list(df[df['University'] == "University of Tehran"]['Category_5'].value_counts()),
    list(df[df['University'] == "University of Tehran"]['Category_6'].value_counts()),
    list(df[df['University'] == "University of Tehran"]['Category_7'].value_counts())
]
y = [
    list(df[df['University'] == "University of Tehran"]['Category_1'].value_counts()),
    list(df[df['University'] == "University of Tehran"]['Category_2'].value_counts()),
    list(df[df['University'] == "University of Tehran"]['Category_3'].value_counts()),
    list(df[df['University'] == "University of Tehran"]['Category_4'].value_counts()),
    list(df[df['University'] == "University of Tehran"]['Category_5'].value_counts()),
    list(df[df['University'] == "University of Tehran"]['Category_6'].value_counts()),
    list(df[df['University'] == "University of Tehran"]['Category_7'].value_counts())
]
sns.barplot(ax=axes[0, 0], x=x[0], y=y[0])
axes[0, 0].set_title("Number of First Categories used in Articles of University of")
axes[0, 0].bar_label(axes[0, 0].containers[0], rotation=90)
axes[0, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 1], x=x[1], y=y[1])
axes[0, 1].set_title("Number of Second Categories used in Articles of University of")
axes[0, 1].bar_label(axes[0, 1].containers[0], rotation=90)
axes[0, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[0, 2], x=x[2], y=y[2])
axes[0, 2].set_title("Number of Third Categories used in Articles of University of")
axes[0, 2].bar_label(axes[0, 2].containers[0], rotation=90)
axes[0, 2].tick_params(labelrotation=90)

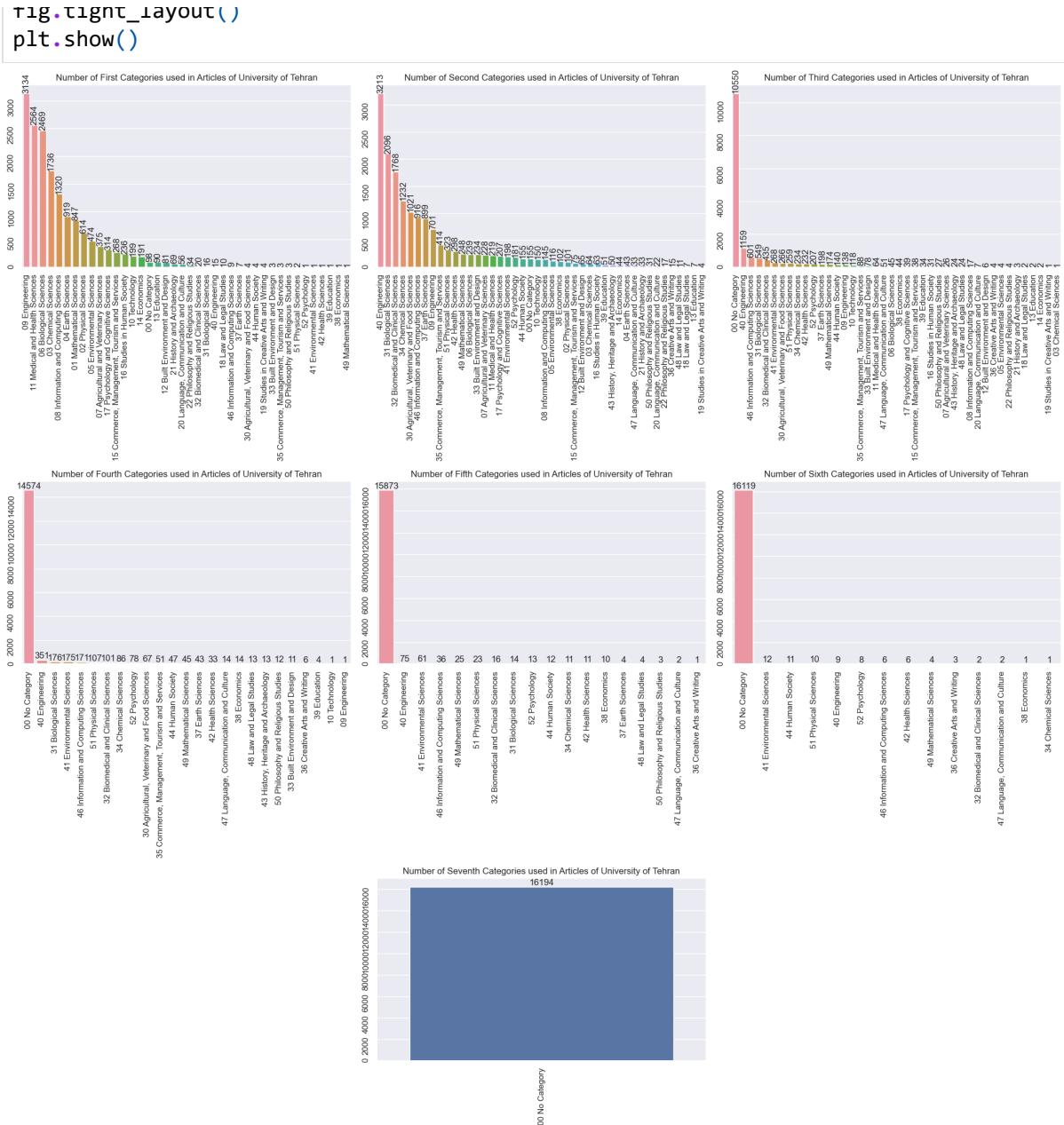
sns.barplot(ax=axes[1, 0], x=x[3], y=y[3])
axes[1, 0].set_title("Number of Fourth Categories used in Articles of University of")
axes[1, 0].bar_label(axes[1, 0].containers[0])
axes[1, 0].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 1], x=x[4], y=y[4])
axes[1, 1].set_title("Number of Fifth Categories used in Articles of University of")
axes[1, 1].bar_label(axes[1, 1].containers[0])
axes[1, 1].tick_params(labelrotation=90)

sns.barplot(ax=axes[1, 2], x=x[5], y=y[5])
axes[1, 2].set_title("Number of Sixth Categories used in Articles of University of")
axes[1, 2].bar_label(axes[1, 2].containers[0])
axes[1, 2].tick_params(labelrotation=90)

sns.barplot(ax=axes[2, 1], x=x[6], y=y[6])
axes[2, 1].set_title("Number of Seventh Categories used in Articles of University o")
axes[2, 1].bar_label(axes[2, 1].containers[0])
axes[2, 1].tick_params(labelrotation=90)

fig.delaxes(axes[2, 0])
fig.delaxes(axes[2, 2])
```



Research Output Type and Open Access situation

- Most of the research outputs tracked in this dataset (92.2%) is article, 7.6% is chapter, 98 books and 1 news also monitored.
 - Most of the published research outputs (65%) is not under open access policies.
 - Out of the outputs published under open access policies, 44% of them published in *Gold* open access journals, 37.4% in *Green* open access journals, 10.2% in *Bronze* open access journals and 7.6% in *Hybrid* open access journals.

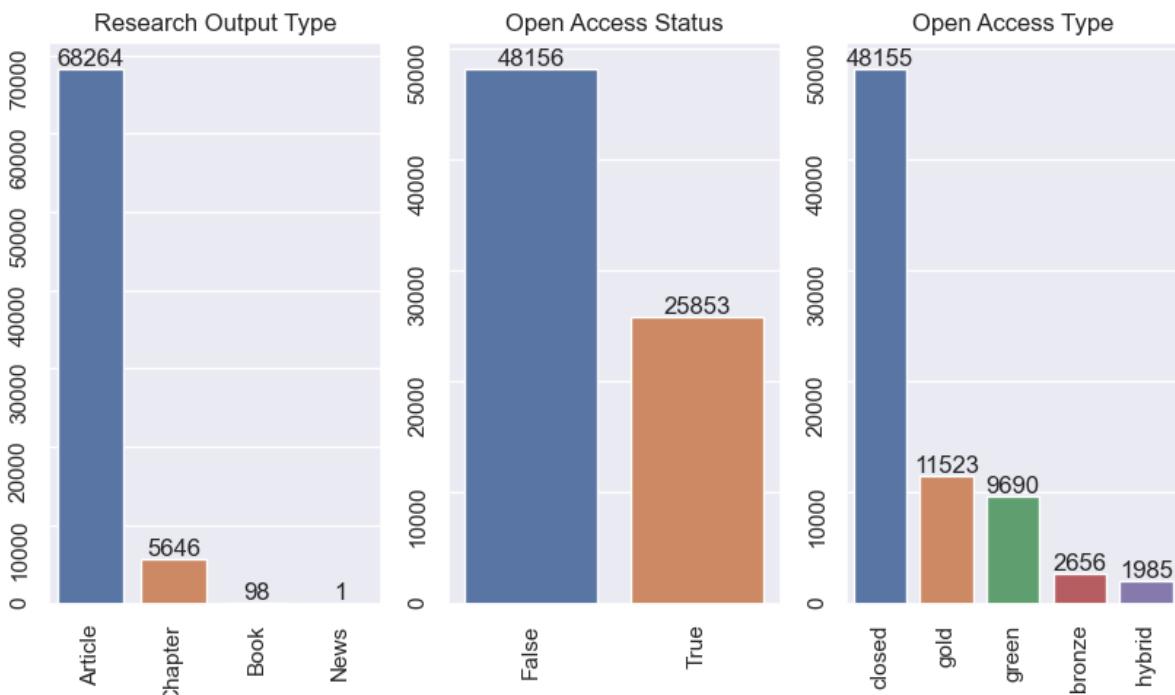
```
In [ ]: fig, axes = plt.subplots(1, 3, figsize=(10, 5))
x = [
    list(df['Output_Type'].value_counts().index),
    list(df['OA_Status'].value_counts().index),
    list(df['OA_Type'].value_counts().index),
]
y = [
    list(df['Output_Type'].value_counts()),
    list(df['OA_Status'].value_counts()),
    list(df['OA_Type'].value_counts()),
]

sns.barplot(ax=axes[0], x=x[0], y=y[0])
axes[0].set_title('Research Output Type')
axes[0].bar_label(axes[0].containers[0])
axes[0].tick_params(labelrotation=90)

sns.barplot(ax=axes[1], x=x[1], y=y[1])
axes[1].set_title('Open Access Status')
axes[1].bar_label(axes[1].containers[0])
axes[1].tick_params(labelrotation=90)

sns.barplot(ax=axes[2], x=x[2], y=y[2])
axes[2].set_title('Open Access Type')
axes[2].bar_label(axes[2].containers[0])
axes[2].tick_params(labelrotation=90)

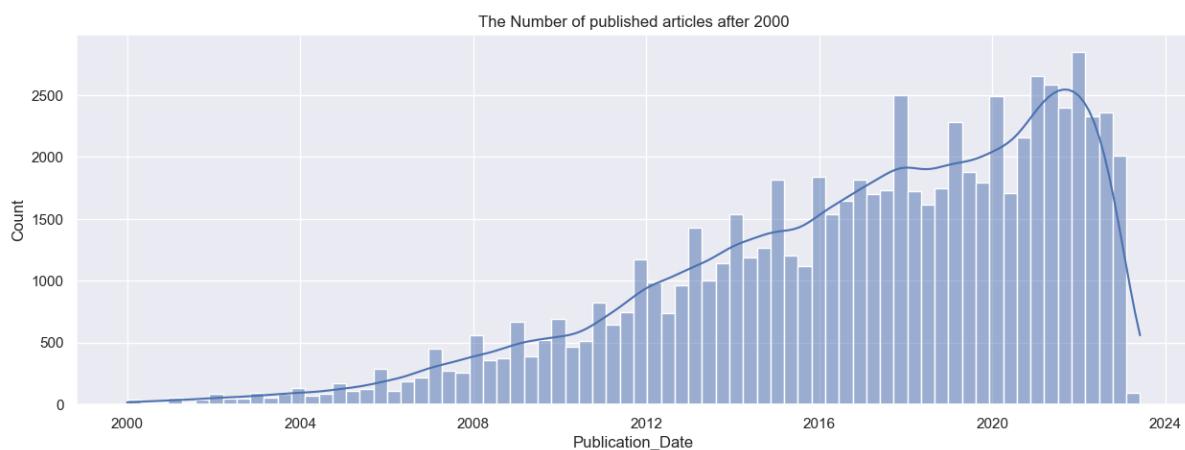
plt.show()
```



Publication Trend

In the plot below you can see the trend of publication based on the date. It's obvious that the trend is growing and more research output is generated. This trait might be due to the increase of tracking research outputs of investigated universities.

```
In [ ]: plt.figure(figsize=(15, 5))
ax = sns.histplot(data=df.loc[df['Publication_Date'] >= '2000'], x='Publication_Dat
ax.bar_label(ax.containers[0])
plt.title('The Number of published articles after 2000')
plt.show()
```



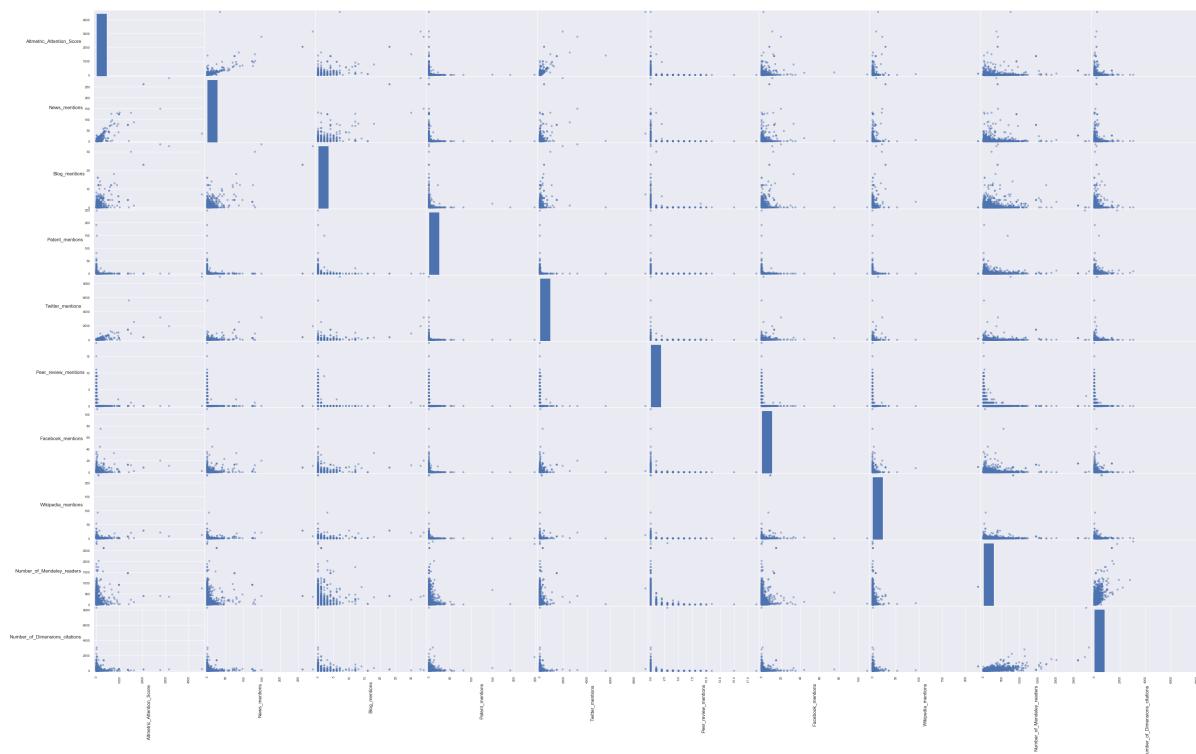
In the next cell you can see the statistical description of each features.

```
In [ ]: df.describe()
```

	Altmetric_Attention_Score	News_mentions	Blog_mentions	Policy_mentions	Patent_mention
count	74009.000000	74009.000000	74009.000000	74009.000000	74009.000000
mean	3.067195	0.138929	0.039576	0.018228	0.161871
std	33.615673	2.703754	0.400570	0.190594	1.664764
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000
50%	1.000000	0.000000	0.000000	0.000000	0.000000
75%	2.000000	0.000000	0.000000	0.000000	0.000000
max	4568.000000	290.000000	34.000000	12.000000	248.000000

In the next cell we are going to visualize a scatter plot for each feature pair. Since we have a good amount of features in the dataset, this plot will be huge. To mitigate this problem, we will remove features with less than 0.25 standard deviation. The Standard Deviation Values are presented in the table above.

```
In [ ]: target_features = list(df.describe().columns[:-13])
temp = []
for feature in target_features:
    if df.describe()[feature]['std'] >= 0.25:
        temp.append(feature)
target_features = temp
axes = pd.plotting.scatter_matrix(df[target_features], figsize=(50, 30), s=100)
for ax in axes.flatten():
    ax.xaxis.label.set_rotation(90)
    ax.yaxis.label.set_rotation(0)
    ax.yaxis.label.set_ha('right')
plt.show()
```

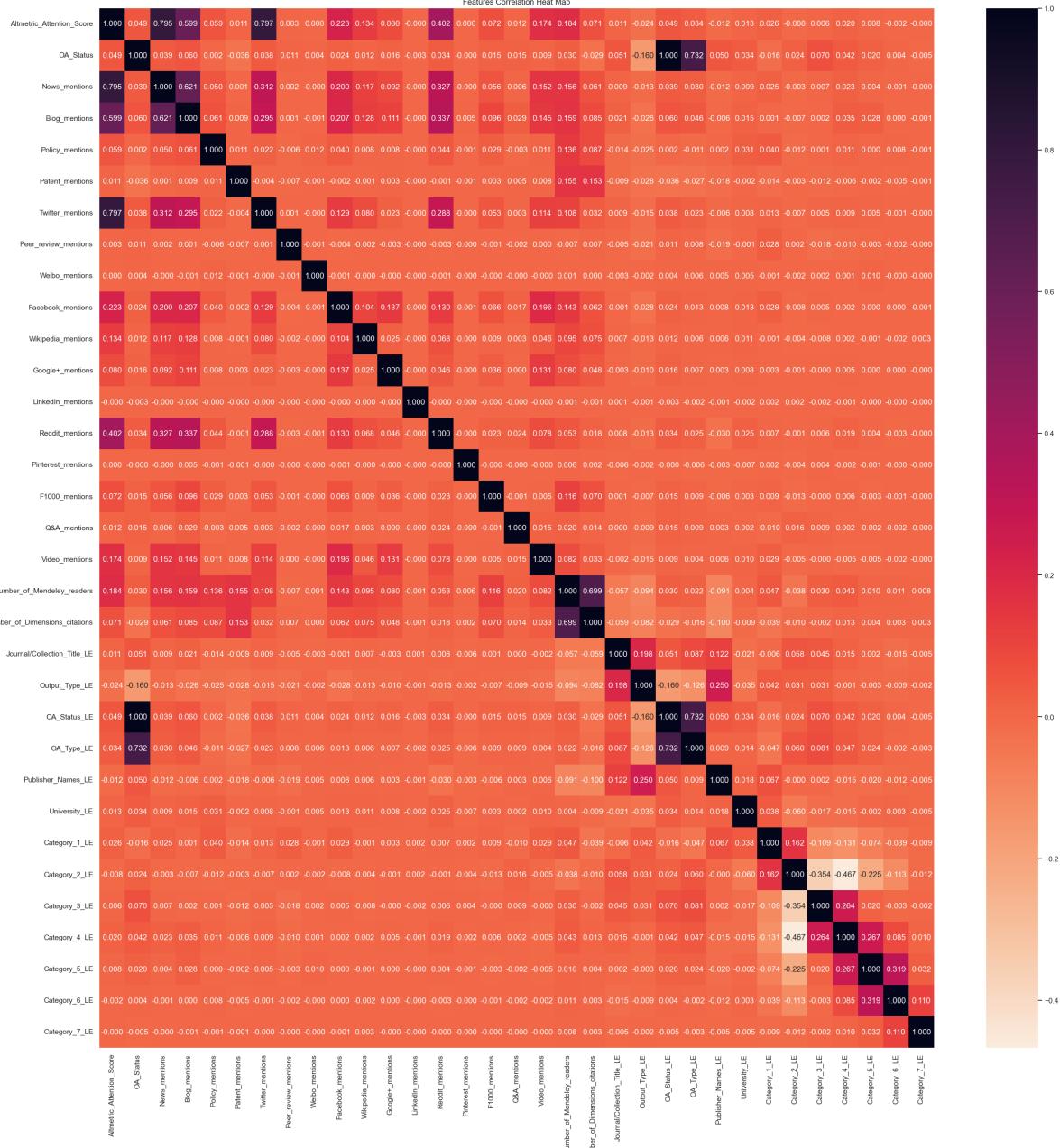


Correlation Analysis

In the next cell we try to explore and analyze the correlation status of features in the dataframe. Have in mind that correlation is meaningless for categorical values, thus we have to make these type of values label encoded. have in mind we will only label encode values that we believe will result in meaningful result.

- There is strong correlation between the **Altmetric Attention Score**, and **News Mentions**, **Blog Mentions** and **Twitter Mentions**. This trait should be further investigated, since we saw on the described statistical features of columns that these features attract the most of the mentions among all type of mentions.
- There are no significance correlation between features, except the two features of Open Access type and status since they are perfectly related.

```
In [ ]: intercor = df.corr(numeric_only=True)
plt.figure(figsize=(25,25))
sns.heatmap(intercor, annot=True, cmap='rocket_r', fmt='.3f')
plt.tight_layout()
plt.title('Features Correlation Heat Map')
plt.show()
```



Grouping Features Analysis

In the cells below we will explore the data by grouping features and calculate the aggregations.

```
In [ ]: df.groupby('Output_Type')[['Altmetric_Attention_Score']].agg(['count', 'mean', 'max',
```

		count	mean	max	min	std
Output_Type						
Article	68264	3.302854	4568	0	34.969362	
Book	98	1.204082	27	0	3.220294	
Chapter	5646	0.246192	220	0	4.284104	
News	1	26.000000	26	26		Nan

```
In [ ]: df.groupby('OA_Status')[['Altmetric_Attention_Score']].agg(['count', 'mean', 'max', 'mi
```

	count	mean	max	min	std	
OA_Status						
False	48156	1.855117	790	0	9.528782	
True	25853	5.324914	4568	0	55.299093	

An interesting insight that we can deduce from the table above, is that publishing under open access policies will result in better performing and more mentions. Although the number of articles under open access policies are lower than closed ones, they resulted in more mentions on average than the closed ones.

```
In [ ]: df.groupby('OA_Type')[['Altmetric_Attention_Score']].agg(['count', 'mean', 'max', 'mi
```

	count	mean	max	min	std	
OA_Type						
bronze	2656	9.338479	4568	0	116.910033	
closed	48155	1.855155	790	0	9.528877	
gold	11523	4.090167	1410	0	22.571000	
green	9690	4.310630	2766	0	47.675711	
hybrid	1985	12.071033	1502	0	86.116285	

```
In [ ]: df.groupby('Publisher_Names')[['Altmetric_Attention_Score']].agg(['count', 'mean', 'mi
```

Out[]:

		count	mean	max	min	std
Publisher_Names						
	Elsevier BV	21585	2.279129	696	0	11.427640
	Springer Nature	13776	1.528818	1017	0	10.665141
	Institute of Electrical and Electronics Engineers (IEEE)	4148	1.224446	95	0	3.094993
	ESSOAr; Natural History Museum; Wiley	3991	2.306439	426	0	11.489564
	GeoScienceWorld; Taylor & Francis	3633	2.218277	282	0	10.420666

	Oxford University Press (OUP); Taylor & Francis	1	0.000000	0	0	NaN
	American Diabetes Association	1	6.000000	6	6	NaN
	Hindawi Limited; Springer Nature	1	3.000000	3	3	NaN
	Consortium Erudit; GeoScienceWorld; Taylor & Francis	1	3.000000	3	3	NaN
	American Thoracic Society; StatRef	1	1.000000	1	1	NaN

189 rows × 5 columns

In []: df.groupby('University')['Altmetric_Attention_Score'].agg(['count', 'mean', 'max',

Out[]:

		count	mean	max	min	std
University						
	University of Tehran	16194	3.995863	4568	0	56.224681
	Tarbiat Modares University	8870	3.917249	2043	0	35.927520
	Sharif University of Technology	6866	2.918730	1367	0	23.913502
	Amir Kabir University	5753	1.788632	1367	0	22.824887
	University of Shiraz	5572	2.689698	633	0	15.228685
	Ferdowsi University of Mashhad	4910	3.114053	1017	0	21.330601
	Shahid Beheshti University	4793	3.412477	1410	0	27.233875
	Isfahan University of Technology	4724	2.270957	372	0	9.669328
	Iran University of Science and Technology	4356	1.435262	507	0	9.761763
	University of Tabriz	4130	1.948184	349	0	8.935398
	University of Isfahan	3103	3.471157	623	0	19.017597
	K. N. Toosi University of Technology	2103	1.725630	221	0	7.409996
	University of Guilan	2049	3.160566	790	0	21.222986
	Allameh Tabataba'i University	461	9.624729	1629	0	88.069070
	Imam Hossein University	125	1.944000	25	0	3.046161

It's worth to mention that, Although Allameh Tabataba'i University has the second to last amount of research output, it has the highest amount of Altmetric attention score on average. It's standard deviation is also highest among the universities. This fact indicates an outlier among the publications of this university.

```
In [ ]: df.groupby('Category_1')['Altmetric_Attention_Score'].agg(['count', 'mean', 'max',
```

Out[]:

		count	mean	max	min	std
Category_1						
	09 Engineering	16476	1.444829	661	0	8.146969
	11 Medical and Health Sciences	10732	6.795471	4568	0	61.368467
	03 Chemical Sciences	10042	1.832503	201	0	4.575673
	06 Biological Sciences	8161	4.476045	2766	0	42.644529
	08 Information and Computing Sciences	6783	1.423411	1017	0	13.187518
	01 Mathematical Sciences	5183	1.368705	190	0	5.173754
	02 Physical Sciences	4605	2.786102	329	0	10.989553
	04 Earth Sciences	2807	4.155682	3149	0	61.895835
	05 Environmental Sciences	1785	4.882913	961	0	33.274107
	07 Agricultural and Veterinary Sciences	1213	1.930750	202	0	6.806992
	17 Psychology and Cognitive Sciences	1167	5.036847	333	0	20.024843
	10 Technology	1146	1.419721	140	0	5.251002
	15 Commerce, Management, Tourism and Services	699	1.589413	141	0	7.767499
	14 Economics	636	2.962264	409	0	19.583904
	16 Studies in Human Society	590	3.288136	175	0	11.463755
	13 Education	427	2.405152	66	0	6.048107
	00 No Category	383	0.950392	69	0	4.294785
	12 Built Environment and Design	333	1.531532	36	0	3.895054
	20 Language, Communication and Culture	203	2.004926	31	0	4.133296
	21 History and Archaeology	138	51.557971	2043	0	281.978050
	22 Philosophy and Religious Studies	121	2.776860	60	0	7.356955
	40 Engineering	82	0.792683	10	0	1.810380
	32 Biomedical and Clinical Sciences	56	1.375000	23	0	3.887334
	18 Law and Legal Studies	43	6.627907	56	0	13.569966
	31 Biological Sciences	39	1.923077	16	0	3.055492
	46 Information and Computing Sciences	25	1.040000	11	0	2.406242
	49 Mathematical Sciences	20	0.150000	2	0	0.489360
	34 Chemical Sciences	17	1.764706	8	0	1.953504
	19 Studies in Creative Arts and Writing	16	2.437500	10	0	3.182635
	37 Earth Sciences	14	0.714286	3	0	1.069045
	35 Commerce, Management, Tourism and Services	13	1.153846	3	0	1.143544
	44 Human Society	8	1.750000	3	1	0.886405

		count	mean	max	min	std
Category_1						
30 Agricultural, Veterinary and Food Sciences	8	2.000000	9	0	3.023716	
50 Philosophy and Religious Studies	8	17.250000	66	0	30.103393	
33 Built Environment and Design	6	1.166667	2	0	0.752773	
51 Physical Sciences	6	2.166667	9	0	3.488075	
52 Psychology	5	2.200000	7	1	2.683282	
42 Health Sciences	3	1.000000	1	1	0.000000	
41 Environmental Sciences	3	0.666667	1	0	0.577350	
47 Language, Communication and Culture	2	0.000000	0	0	0.000000	
39 Education	2	2.000000	3	1	1.414214	
38 Economics	2	0.500000	1	0	0.707107	
36 Creative Arts and Writing	1	0.000000	0	0	NaN	

As it was discussed earlier, the majority of research output had **Engineering** as their category. But the majority of attentions was retrieved by the research output which had **Medical and Health Sciences** as their first category. This trait is probably due to the Covid-19 pandemic and the fact that research about this pandemic retrieved much more attention in the media.

Another interesting insight, is the fact that some research outputs in category of **21 History and Archaeology** attract some attention, after further investigation, it was concluded that there is a research output with about 2000 attention score that was both credited to University of Tehran and Tarbiat Modares University.

```
In [ ]: df.groupby('Category_2')['Altmetric_Attention_Score'].agg(['count', 'mean', 'max',
```

Out[]:

		count	mean	max	min	std
Category_2						
	40 Engineering	17490	1.260835	369	0	4.751832
	32 Biomedical and Clinical Sciences	7645	6.060562	4568	0	60.612184
	34 Chemical Sciences	7452	1.785293	182	0	4.002456
	31 Biological Sciences	6882	4.774920	1629	0	36.021628
	46 Information and Computing Sciences	4580	1.340175	1017	0	15.497656
	09 Engineering	3873	1.709786	191	0	5.868211
	30 Agricultural, Veterinary and Food Sciences	3487	3.680528	2043	0	55.416739
	37 Earth Sciences	2817	4.046858	3149	0	62.551944
	51 Physical Sciences	2771	3.561891	507	0	17.742544
	49 Mathematical Sciences	2163	1.509940	45	0	2.946684
	42 Health Sciences	1416	9.480932	1367	0	70.732720
	35 Commerce, Management, Tourism and Services	1243	1.448914	179	0	8.003598
	11 Medical and Health Sciences	1061	3.835061	948	0	29.741752
	06 Biological Sciences	939	6.604899	961	0	43.754115
	33 Built Environment and Design	895	2.879330	661	0	25.183465
	02 Physical Sciences	826	2.829298	102	0	8.106483
	10 Technology	771	1.688716	28	0	2.876978
	41 Environmental Sciences	764	3.738220	897	0	33.659922
	07 Agricultural and Veterinary Sciences	738	1.897019	49	0	3.412837
	17 Psychology and Cognitive Sciences	734	6.359673	585	0	32.664048
	08 Information and Computing Sciences	677	1.109306	37	0	2.475296
	52 Psychology	654	7.568807	471	0	32.072668
	00 No Category	582	1.130584	69	0	3.879439
	05 Environmental Sciences	397	2.617128	215	0	11.783806
	44 Human Society	375	3.765333	111	0	10.423563
	03 Chemical Sciences	342	3.017544	134	0	10.337685
	38 Economics	319	2.344828	119	0	9.500425
	39 Education	317	2.372240	44	0	5.253713
	12 Built Environment and Design	292	3.071918	243	0	17.173812
	15 Commerce, Management, Tourism and Services	260	1.334615	61	0	5.533293
	16 Studies in Human Society	178	3.949438	66	0	9.710075
	14 Economics	174	1.689655	31	0	3.868684

		count	mean	max	min	std
Category_2						
47 Language, Communication and Culture		163	2.036810	24	0	3.656365
04 Earth Sciences		125	2.120000	16	0	2.894934
50 Philosophy and Religious Studies		117	3.470085	60	0	8.375429
43 History, Heritage and Archaeology		104	5.644231	88	0	14.084294
20 Language, Communication and Culture		103	1.611650	22	0	2.762611
21 History and Archaeology		60	91.933333	2766	0	374.473681
36 Creative Arts and Writing		48	2.604167	31	0	4.832248
48 Law and Legal Studies		41	6.731707	56	0	13.881686
22 Philosophy and Religious Studies		38	6.657895	175	0	28.223314
13 Education		37	2.000000	13	0	3.291403
18 Law and Legal Studies		31	1.741935	9	0	2.780500
19 Studies in Creative Arts and Writing		28	1.642857	8	0	1.591977

In []: df.groupby('Category_3')[['Altmetric_Attention_Score']].agg(['count', 'mean', 'max',

Out[]:

		count	mean	max	min	std
Category_3						
	00 No Category	48454	2.858773	4568	0	31.506775
	40 Engineering	5978	1.650719	201	0	5.741645
	46 Information and Computing Sciences	2977	1.425596	357	0	8.054330
	51 Physical Sciences	2184	2.132784	173	0	6.222088
	31 Biological Sciences	1932	10.700311	2766	0	106.481177
	32 Biomedical and Clinical Sciences	1753	5.353679	1629	0	48.307701
	34 Chemical Sciences	1293	2.003094	134	0	5.454852
	49 Mathematical Sciences	1134	1.514109	59	0	4.306369
	42 Health Sciences	976	8.141393	409	0	28.374031
	30 Agricultural, Veterinary and Food Sciences	950	1.887368	110	0	4.892625
	41 Environmental Sciences	947	3.059134	430	0	19.006249
	09 Engineering	678	2.246313	102	0	8.314699
	52 Psychology	673	7.595840	288	0	26.560243
	37 Earth Sciences	653	2.540582	215	0	10.938911
	10 Technology	611	2.217676	191	0	8.242836
	44 Human Society	387	14.953488	1367	0	120.638856
	35 Commerce, Management, Tourism and Services	301	1.631229	61	0	6.369737
	33 Built Environment and Design	278	2.672662	243	0	14.945534
	11 Medical and Health Sciences	275	3.047273	191	0	11.871327
	06 Biological Sciences	175	2.411429	55	0	6.605630
	38 Economics	173	5.086705	661	0	50.224363
	47 Language, Communication and Culture	166	2.584337	42	0	4.666008
	15 Commerce, Management, Tourism and Services	154	1.461039	109	0	8.872468
	17 Psychology and Cognitive Sciences	133	1.827068	44	0	4.484842
	07 Agricultural and Veterinary Sciences	111	1.270270	17	0	2.276145
	39 Education	110	1.854545	15	0	2.573101
	16 Studies in Human Society	104	2.528846	31	0	5.307911
	50 Philosophy and Religious Studies	83	3.807229	175	0	19.158809
	08 Information and Computing Sciences	74	2.175676	13	0	2.953327
	48 Law and Legal Studies	72	2.777778	52	0	7.027556
	43 History, Heritage and Archaeology	55	26.854545	792	0	115.114342
	20 Language, Communication and Culture	34	0.970588	8	0	1.660327

		count	mean	max	min	std
Category_3						
	12 Built Environment and Design	25	1.960000	12	0	3.034249
	03 Chemical Sciences	22	0.772727	3	0	0.869144
	36 Creative Arts and Writing	15	1.066667	8	0	2.086236
	22 Philosophy and Religious Studies	13	2.076923	9	0	2.531848
	05 Environmental Sciences	13	0.615385	2	0	0.650444
	18 Law and Legal Studies	12	1.750000	4	0	1.138180
	14 Economics	10	0.900000	6	0	1.852926
	19 Studies in Creative Arts and Writing	10	2.000000	6	0	1.943651
	13 Education	6	5.333333	19	0	7.201852
	21 History and Archaeology	5	1.800000	6	0	2.387467

```
In [ ]: df.groupby('Category_4')[['Altmetric_Attention_Score']].agg(['count', 'mean', 'max',
```

Out[]:

		count	mean	max	min	std
Category_4						
	00 No Category	66832	2.879938	4568	0	30.058128
	40 Engineering	1883	1.771110	82	0	3.946935
	46 Information and Computing Sciences	797	1.908407	227	0	13.306888
	51 Physical Sciences	785	2.719745	74	0	6.716408
	41 Environmental Sciences	643	9.191291	961	0	58.319715
	31 Biological Sciences	616	2.217532	110	0	5.558974
	32 Biomedical and Clinical Sciences	414	2.748792	63	0	5.567954
	34 Chemical Sciences	316	2.984177	191	0	12.253787
	49 Mathematical Sciences	290	2.000000	102	0	10.691005
	52 Psychology	272	9.904412	585	0	48.404870
	30 Agricultural, Veterinary and Food Sciences	222	2.072072	54	0	4.724465
	35 Commerce, Management, Tourism and Services	179	1.636872	109	0	8.332081
	37 Earth Sciences	164	3.756098	241	0	19.546742
	44 Human Society	154	9.928571	792	0	66.844773
	42 Health Sciences	129	4.449612	59	0	8.645773
	47 Language, Communication and Culture	69	3.855072	66	0	11.210913
	33 Built Environment and Design	59	1.101695	11	0	2.179014
	38 Economics	50	3.260000	108	0	15.168186
	48 Law and Legal Studies	37	4.837838	108	0	17.659611
	36 Creative Arts and Writing	29	1.931034	8	0	1.869512
	43 History, Heritage and Archaeology	24	442.541667	2766	0	797.265707
	50 Philosophy and Religious Studies	21	1.571429	15	0	3.264528
	39 Education	18	2.277778	12	0	3.922867
	10 Technology	3	0.666667	1	0	0.577350
	09 Engineering	2	1.000000	1	1	0.000000
	11 Medical and Health Sciences	1	6.000000	6	6	NaN

In []: df.groupby('Category_5')[['Altmetric_Attention_Score']].agg(['count', 'mean', 'max',

Out[]:

		count	mean	max	min	std
Category_5						
	00 No Category	72667	3.037761	4568	0	32.307099
	40 Engineering	331	2.700906	191	0	11.599816
	41 Environmental Sciences	204	2.049020	54	0	5.236058
	46 Information and Computing Sciences	177	1.468927	23	0	2.569576
	51 Physical Sciences	145	4.172414	102	0	14.840516
	49 Mathematical Sciences	95	1.684211	21	0	3.498520
	34 Chemical Sciences	82	2.036585	24	0	3.469248
	32 Biomedical and Clinical Sciences	64	1.703125	11	0	2.044580
	52 Psychology	41	4.780488	27	0	5.667946
	44 Human Society	38	75.605263	2766	0	448.337924
	31 Biological Sciences	36	1.055556	5	0	1.119807
	42 Health Sciences	33	4.151515	44	0	8.039439
	38 Economics	24	2.416667	11	0	3.525271
	47 Language, Communication and Culture	17	1.235294	8	0	2.136861
	37 Earth Sciences	13	2.000000	6	0	2.198484
	35 Commerce, Management, Tourism and Services	12	0.333333	2	0	0.651339
	48 Law and Legal Studies	10	1.000000	3	0	1.154701
	50 Philosophy and Religious Studies	9	1.000000	5	0	1.581139
	33 Built Environment and Design	4	1.750000	6	0	2.872281
	43 History, Heritage and Archaeology	4	64.500000	241	0	117.854430
	36 Creative Arts and Writing	3	1.666667	5	0	2.886751

In []: df.groupby('Category_6')[['Altmetric_Attention_Score']].agg(['count', 'mean', 'max',

Out[]:

		count	mean	max	min	std
Category_6						
00 No Category	73746	3.070865	4568	0	33.674597	
41 Environmental Sciences	61	2.180328	24	0	3.761685	
51 Physical Sciences	41	2.536585	15	0	3.795376	
44 Human Society	32	2.750000	22	0	4.898979	
40 Engineering	27	1.296296	10	0	2.127021	
46 Information and Computing Sciences	27	1.518519	10	0	2.562606	
49 Mathematical Sciences	22	0.454545	2	0	0.670982	
52 Psychology	19	3.789474	44	0	10.003216	
42 Health Sciences	12	1.416667	4	1	0.900337	
36 Creative Arts and Writing	9	2.555556	6	0	2.297341	
32 Biomedical and Clinical Sciences	4	2.000000	3	1	0.816497	
47 Language, Communication and Culture	4	0.500000	1	0	0.577350	
34 Chemical Sciences	3	0.333333	1	0	0.577350	
38 Economics	2	1.000000	1	1	0.000000	

In []: df.groupby('Category_7')[['Altmetric_Attention_Score']].agg(['count', 'mean', 'max', 'min', 'std'])

Out[]:

	count	mean	max	min	std
Category_7					
00 No Category	74006	3.067238	4568	0	33.616353
49 Mathematical Sciences	3	2.000000	3	0	1.732051

Group by Top Category for each University

Allameh Tabataba'i University

In []: df[df['University'] == "Allameh Tabataba'i University"].groupby('Category_1')[['Altmetric_Attention_Score']].agg(['count', 'mean', 'max', 'min', 'std'])

Out[]:

		count	mean	max	min	std
Category_1						
	11 Medical and Health Sciences	116	29.724138	1629	0	173.372511
	17 Psychology and Cognitive Sciences	58	3.172414	31	0	5.959330
	08 Information and Computing Sciences	45	1.311111	13	0	2.475537
	01 Mathematical Sciences	39	1.717949	40	0	6.353402
	15 Commerce, Management, Tourism and Services	39	3.435897	97	0	15.449059
	13 Education	34	3.088235	66	0	11.163706
	16 Studies in Human Society	29	7.965517	175	0	32.206347
	09 Engineering	20	0.950000	4	0	1.276302
	14 Economics	20	2.700000	16	0	4.612340
	20 Language, Communication and Culture	19	2.578947	24	0	5.610662
	12 Built Environment and Design	7	1.000000	3	0	1.000000
	03 Chemical Sciences	6	1.000000	3	0	1.264911
	21 History and Archaeology	5	1.200000	4	0	1.643168
	22 Philosophy and Religious Studies	4	1.000000	3	0	1.414214
	06 Biological Sciences	4	2.250000	6	0	2.629956
	18 Law and Legal Studies	4	7.750000	18	0	7.500000
	00 No Category	3	0.000000	0	0	0.000000
	10 Technology	3	4.000000	9	0	4.582576
	19 Studies in Creative Arts and Writing	1	0.000000	0	0	NaN
	07 Agricultural and Veterinary Sciences	1	3.000000	3	3	NaN
	05 Environmental Sciences	1	4.000000	4	4	NaN
	02 Physical Sciences	1	3.000000	3	3	NaN
	35 Commerce, Management, Tourism and Services	1	1.000000	1	1	NaN
	46 Information and Computing Sciences	1	1.000000	1	1	NaN

Amir Kabir University

In []: df[df['University'] == "Amir Kabir University"].groupby('Category_1')[['Altmetric_Av

Out[]:

		count	mean	max	min	std
Category_1						
	09 Engineering	2208	1.057518	28	0	1.986554
	08 Information and Computing Sciences	816	1.182598	104	0	4.861744
	03 Chemical Sciences	791	1.656131	106	0	4.291951
	01 Mathematical Sciences	583	0.910806	59	0	2.841567
	11 Medical and Health Sciences	390	9.964103	1367	0	86.568716
	02 Physical Sciences	257	1.466926	12	0	1.932422
	06 Biological Sciences	137	1.956204	34	0	3.886002
	04 Earth Sciences	132	1.416667	26	0	2.887302
	10 Technology	117	0.923077	7	0	1.457169
	00 No Category	60	0.316667	6	0	1.065510
15 Commerce, Management, Tourism and Services		50	0.580000	5	0	1.070762
	05 Environmental Sciences	43	1.511628	22	0	3.500830
	17 Psychology and Cognitive Sciences	41	1.292683	7	0	1.569138
	14 Economics	31	1.096774	9	0	2.134711
	12 Built Environment and Design	19	1.157895	3	0	1.067872
	16 Studies in Human Society	17	1.411765	15	0	3.742640
	13 Education	12	1.666667	3	0	1.073087
	22 Philosophy and Religious Studies	9	0.888889	3	0	0.927961
	32 Biomedical and Clinical Sciences	7	1.142857	8	0	3.023716
	40 Engineering	6	0.000000	0	0	0.000000
	07 Agricultural and Veterinary Sciences	6	2.500000	9	0	3.728270
	49 Mathematical Sciences	4	0.500000	2	0	1.000000
	21 History and Archaeology	3	0.000000	0	0	0.000000
35 Commerce, Management, Tourism and Services		3	0.000000	0	0	0.000000
	46 Information and Computing Sciences	3	3.666667	11	0	6.350853
	20 Language, Communication and Culture	2	1.500000	3	0	2.121320
	18 Law and Legal Studies	2	1.000000	1	1	0.000000
	34 Chemical Sciences	1	8.000000	8	8	NaN
	37 Earth Sciences	1	0.000000	0	0	NaN
	47 Language, Communication and Culture	1	0.000000	0	0	NaN
	50 Philosophy and Religious Studies	1	0.000000	0	0	NaN

Shahid Beheshti University

```
In [ ]: df[df['University'] == "Shahid Beheshti University"].groupby('Category_1')['Altmetr
```

Out[]:

		count	mean	max	min	std
Category_1						
	11 Medical and Health Sciences	895	6.423464	764	0	37.688060
	03 Chemical Sciences	661	1.996974	60	0	3.674646
	09 Engineering	600	1.428333	97	0	4.800688
	02 Physical Sciences	566	2.994700	219	0	11.449233
	06 Biological Sciences	496	6.508065	1410	0	64.434008
	08 Information and Computing Sciences	379	1.897098	191	0	10.525226
	01 Mathematical Sciences	343	1.472303	19	0	2.475309
	17 Psychology and Cognitive Sciences	169	5.532544	220	0	18.026117
	04 Earth Sciences	165	2.375758	25	0	4.008149
	05 Environmental Sciences	121	2.768595	35	0	4.900273
	10 Technology	76	1.184211	10	0	1.902353
	16 Studies in Human Society	52	2.134615	16	0	3.211675
	07 Agricultural and Veterinary Sciences	44	2.250000	13	0	3.170540
	14 Economics	43	0.906977	7	0	1.394102
	00 No Category	34	2.205882	69	0	11.813653
	13 Education	29	1.655172	9	0	2.334154
	15 Commerce, Management, Tourism and Services	27	1.740741	14	0	2.781707
	12 Built Environment and Design	27	1.555556	22	0	4.181768
	20 Language, Communication and Culture	21	1.714286	18	0	3.887710
	40 Engineering	10	0.400000	3	0	0.966092
	22 Philosophy and Religious Studies	10	0.200000	1	0	0.421637
	21 History and Archaeology	7	1.571429	3	0	1.272418
	18 Law and Legal Studies	5	0.800000	3	0	1.303840
	49 Mathematical Sciences	3	0.333333	1	0	0.577350
	31 Biological Sciences	2	0.500000	1	0	0.707107
	32 Biomedical and Clinical Sciences	2	2.000000	3	1	1.414214
	33 Built Environment and Design	1	2.000000	2	2	NaN
	34 Chemical Sciences	1	1.000000	1	1	NaN
	42 Health Sciences	1	1.000000	1	1	NaN
	19 Studies in Creative Arts and Writing	1	0.000000	0	0	NaN
	50 Philosophy and Religious Studies	1	1.000000	1	1	NaN
	51 Physical Sciences	1	2.000000	2	2	NaN

Ferdowsi University of Mashhad

```
In [ ]: df[df['University'] == "Ferdowsi University of Mashhad"].groupby('Category_1')['Alt
```

Out[]:

		count	mean	max	min	std
Category_1						
	09 Engineering	913	1.197152	77	0	3.189984
	06 Biological Sciences	705	4.012766	220	0	14.603417
	11 Medical and Health Sciences	657	5.474886	357	0	25.807089
	03 Chemical Sciences	628	1.815287	131	0	5.690667
	08 Information and Computing Sciences	409	3.493888	1017	0	50.272547
	01 Mathematical Sciences	352	0.920455	10	0	1.588575
	02 Physical Sciences	259	2.467181	63	0	4.962835
	04 Earth Sciences	230	4.517391	241	0	19.943964
	05 Environmental Sciences	157	11.280255	538	0	52.470240
	07 Agricultural and Veterinary Sciences	151	1.761589	24	0	3.097760
	17 Psychology and Cognitive Sciences	99	4.676768	139	0	14.819890
	10 Technology	54	1.333333	9	0	2.136829
	15 Commerce, Management, Tourism and Services	52	2.134615	82	0	11.341454
	13 Education	51	1.333333	10	0	1.544884
	14 Economics	44	0.659091	5	0	0.963115
	16 Studies in Human Society	35	2.514286	37	0	6.227885
	20 Language, Communication and Culture	27	2.111111	16	0	3.377907
	12 Built Environment and Design	20	0.900000	3	0	0.967906
	00 No Category	14	4.285714	25	0	8.165863
	22 Philosophy and Religious Studies	12	6.750000	46	0	12.892034
	21 History and Archaeology	9	1.888889	5	0	1.763834
	32 Biomedical and Clinical Sciences	5	9.400000	23	0	9.555103
	40 Engineering	5	0.000000	0	0	0.000000
	19 Studies in Creative Arts and Writing	4	3.500000	8	0	3.696846
	31 Biological Sciences	4	4.250000	11	0	4.716991
	34 Chemical Sciences	2	0.500000	1	0	0.707107
	35 Commerce, Management, Tourism and Services	2	3.000000	3	3	0.000000
	50 Philosophy and Religious Studies	2	1.000000	1	1	0.000000
	37 Earth Sciences	1	3.000000	3	3	NaN
	38 Economics	1	1.000000	1	1	NaN
	39 Education	1	3.000000	3	3	NaN
	30 Agricultural, Veterinary and Food Sciences	1	3.000000	3	3	NaN

		count	mean	max	min	std
Category_1						
	44 Human Society	1	1.000000	1	1	NaN
	46 Information and Computing Sciences	1	0.000000	0	0	NaN
	49 Mathematical Sciences	1	0.000000	0	0	NaN
	52 Psychology	1	1.000000	1	1	NaN

University of Guilan

```
In [ ]: df[df['University'] == "University of Guilan"].groupby('Category_1')['Altmetric_Att
```

Out[]:

		count	mean	max	min	std
Category_1						
	09 Engineering	370	1.621622	100	0	5.938033
	06 Biological Sciences	361	4.277008	295	0	19.403772
	11 Medical and Health Sciences	288	9.291667	790	0	50.066976
	03 Chemical Sciences	282	1.652482	14	0	1.842503
	02 Physical Sciences	231	1.333333	31	0	3.212295
	01 Mathematical Sciences	109	1.055046	13	0	1.819823
	08 Information and Computing Sciences	102	1.166667	22	0	2.552744
	05 Environmental Sciences	74	3.324324	182	0	21.081127
	07 Agricultural and Veterinary Sciences	62	1.483871	10	0	2.078318
	04 Earth Sciences	40	1.150000	12	0	2.019774
	17 Psychology and Cognitive Sciences	25	3.200000	24	1	5.730038
	16 Studies in Human Society	21	1.666667	9	0	1.906130
	10 Technology	18	1.055556	3	0	0.937595
	13 Education	15	3.600000	34	0	8.517209
	15 Commerce, Management, Tourism and Services	11	0.909091	3	0	0.943880
	00 No Category	10	2.400000	17	0	5.253570
	12 Built Environment and Design	7	1.000000	1	1	0.000000
	14 Economics	7	1.000000	3	0	1.000000
	20 Language, Communication and Culture	4	3.500000	4	2	1.000000
	22 Philosophy and Religious Studies	3	1.333333	3	0	1.527525
	31 Biological Sciences	3	1.333333	2	1	0.577350
	18 Law and Legal Studies	2	0.500000	1	0	0.707107
	34 Chemical Sciences	2	2.000000	3	1	1.414214
	21 History and Archaeology	1	0.000000	0	0	NaN
	32 Biomedical and Clinical Sciences	1	1.000000	1	1	NaN

Imam Hossein University

In []: df[df['University'] == "Imam Hossein University"].groupby('Category_1')[['Altmetric_

Out[]:

		count	mean	max	min	std
Category_1						
	03 Chemical Sciences	44	2.090909	10	0	2.009491
	11 Medical and Health Sciences	26	2.846154	25	0	5.334359
	09 Engineering	14	1.000000	6	0	1.797434
	02 Physical Sciences	11	1.272727	6	0	1.618080
	01 Mathematical Sciences	7	0.714286	2	0	0.755929
08 Information and Computing Sciences		7	1.285714	3	0	1.253566
17 Psychology and Cognitive Sciences		6	3.166667	11	0	4.490731
	06 Biological Sciences	4	2.250000	4	1	1.500000
	00 No Category	1	3.000000	3	3	NaN
	04 Earth Sciences	1	0.000000	0	0	NaN
07 Agricultural and Veterinary Sciences		1	0.000000	0	0	NaN
	12 Built Environment and Design	1	1.000000	1	1	NaN
	14 Economics	1	0.000000	0	0	NaN
	34 Chemical Sciences	1	3.000000	3	3	NaN

University of Isfahan

```
In [ ]: df[df['University'] == "University of Isfahan"].groupby('Category 1')['Altmetric At
```

Out[]:

		count	mean	max	min	std
Category_1						
	11 Medical and Health Sciences	793	7.934426	623	0	35.987242
	03 Chemical Sciences	487	1.874743	55	0	3.833832
	09 Engineering	395	1.579747	74	0	4.661713
	06 Biological Sciences	349	2.289398	66	0	4.608766
	08 Information and Computing Sciences	241	1.560166	47	0	3.763519
	01 Mathematical Sciences	173	1.023121	19	0	2.048691
	02 Physical Sciences	166	1.072289	12	0	1.305310
	17 Psychology and Cognitive Sciences	109	4.908257	117	0	15.432974
	04 Earth Sciences	81	2.469136	23	0	4.505792
	15 Commerce, Management, Tourism and Services	60	1.133333	8	0	1.770442
	13 Education	51	2.313725	17	0	3.770890
	10 Technology	34	1.000000	6	0	1.348400
	05 Environmental Sciences	32	5.812500	114	0	20.233217
	16 Studies in Human Society	29	4.241379	64	0	11.897344
	20 Language, Communication and Culture	23	1.826087	9	0	2.405527
	21 History and Archaeology	18	1.555556	4	0	1.041618
	12 Built Environment and Design	14	0.928571	5	0	1.491735
	14 Economics	14	1.357143	5	0	1.736803
	00 No Category	9	1.333333	10	0	3.278719
	07 Agricultural and Veterinary Sciences	8	0.625000	2	0	0.744024
	22 Philosophy and Religious Studies	6	1.166667	3	0	0.983192
	18 Law and Legal Studies	3	4.000000	9	0	4.582576
	31 Biological Sciences	2	1.500000	3	0	2.121320
	44 Human Society	2	2.000000	3	1	1.414214
	30 Agricultural, Veterinary and Food Sciences	1	1.000000	1	1	NaN
	32 Biomedical and Clinical Sciences	1	0.000000	0	0	NaN
	46 Information and Computing Sciences	1	0.000000	0	0	NaN
	51 Physical Sciences	1	2.000000	2	2	NaN

Isfahan University of Technology

In []: df[df['University'] == "Isfahan University of Technology"].groupby('Category_1')['A

Out[]:

		count	mean	max	min	std
Category_1						
	09 Engineering	1287	1.590521	135	0	5.982557
	03 Chemical Sciences	793	1.559899	39	0	2.388262
	02 Physical Sciences	567	4.128748	238	0	12.154744
	01 Mathematical Sciences	449	2.821826	190	0	11.233296
	06 Biological Sciences	415	2.392771	86	0	6.073817
	08 Information and Computing Sciences	412	0.720874	10	0	1.435474
	05 Environmental Sciences	215	2.827907	66	0	8.515141
	11 Medical and Health Sciences	148	4.337838	98	0	12.035409
	04 Earth Sciences	144	6.236111	372	0	36.190635
	07 Agricultural and Veterinary Sciences	117	1.649573	11	0	2.229567
	10 Technology	60	1.166667	14	0	2.293444
	00 No Category	53	0.264151	6	0	1.162738
	12 Built Environment and Design	12	0.833333	1	0	0.389249
	14 Economics	11	0.636364	3	0	1.026911
	16 Studies in Human Society	9	0.888889	3	0	1.054093
	17 Psychology and Cognitive Sciences	7	4.285714	21	0	7.653197
	13 Education	5	0.600000	2	0	0.894427
	15 Commerce, Management, Tourism and Services	5	1.200000	3	0	1.095445
	32 Biomedical and Clinical Sciences	4	0.000000	0	0	0.000000
	40 Engineering	4	0.250000	1	0	0.500000
	18 Law and Legal Studies	3	16.000000	44	0	24.331050
	30 Agricultural, Veterinary and Food Sciences	1	0.000000	0	0	NaN
	31 Biological Sciences	1	0.000000	0	0	NaN
	33 Built Environment and Design	1	1.000000	1	1	NaN
	52 Psychology	1	7.000000	7	7	NaN

K. N. Toosi University of Technology

In []: df[df['University'] == "K. N. Toosi University of Technology"].groupby('Category_1')

Out[]:

		count	mean	max	min	std
Category_1						
	09 Engineering	752	1.285904	63	0	3.891253
	08 Information and Computing Sciences	325	1.363077	57	0	3.650282
	03 Chemical Sciences	284	1.679577	17	0	2.172622
	01 Mathematical Sciences	188	0.851064	11	0	1.447531
	11 Medical and Health Sciences	123	3.943089	221	0	19.980031
	02 Physical Sciences	120	1.816667	14	0	2.425299
	04 Earth Sciences	112	1.732143	12	0	2.471394
	10 Technology	58	1.879310	10	0	2.421233
	06 Biological Sciences	36	2.055556	11	0	3.134549
	05 Environmental Sciences	30	2.733333	17	0	4.101584
	14 Economics	15	23.000000	179	0	52.607713
	17 Psychology and Cognitive Sciences	10	1.200000	3	0	1.316561
	12 Built Environment and Design	9	0.888889	2	0	0.781736
	15 Commerce, Management, Tourism and Services	9	2.555556	12	0	3.711843
	16 Studies in Human Society	8	0.625000	2	0	0.744024
	00 No Category	5	0.200000	1	0	0.447214
	07 Agricultural and Veterinary Sciences	5	1.000000	3	0	1.224745
	18 Law and Legal Studies	3	1.333333	4	0	2.309401
	40 Engineering	3	4.000000	6	0	3.464102
	34 Chemical Sciences	2	0.500000	1	0	0.707107
	19 Studies in Creative Arts and Writing	1	1.000000	1	1	NaN
	20 Language, Communication and Culture	1	0.000000	0	0	NaN
	21 History and Archaeology	1	1.000000	1	1	NaN
	22 Philosophy and Religious Studies	1	0.000000	0	0	NaN
	31 Biological Sciences	1	2.000000	2	2	NaN
	32 Biomedical and Clinical Sciences	1	0.000000	0	0	NaN

Sharif University of Technology

In []: df[df['University'] == "Sharif University of Technology"].groupby('Category_1')[['A1

Out[]:

		count	mean	max	min	std
Category_1						
	09 Engineering	2001	1.469765	140	0	5.305866
	03 Chemical Sciences	1029	2.423712	201	0	8.331604
	08 Information and Computing Sciences	1014	1.288955	108	0	4.367647
	01 Mathematical Sciences	799	1.630788	102	0	6.030424
	02 Physical Sciences	765	4.913725	329	0	21.052363
	11 Medical and Health Sciences	431	13.236659	1367	0	85.964972
	10 Technology	194	1.788660	140	0	10.129814
	06 Biological Sciences	161	5.875776	304	0	24.832126
	04 Earth Sciences	97	4.979381	126	0	17.562970
	14 Economics	65	1.600000	14	0	2.793967
	17 Psychology and Cognitive Sciences	59	1.338983	19	0	3.014964
	15 Commerce, Management, Tourism and Services	42	4.642857	141	0	21.725361
	05 Environmental Sciences	36	1.694444	14	0	2.925938
	00 No Category	30	0.900000	15	0	2.795933
	40 Engineering	22	0.772727	6	0	1.823963
	16 Studies in Human Society	20	2.100000	10	0	2.900091
	13 Education	17	1.823529	7	0	2.038237
	12 Built Environment and Design	15	1.466667	6	0	1.684665
	22 Philosophy and Religious Studies	14	7.214286	60	0	15.870527
	20 Language, Communication and Culture	13	1.769231	6	0	2.087816
	49 Mathematical Sciences	8	0.000000	0	0	0.000000
	34 Chemical Sciences	6	1.666667	3	0	1.211060
	18 Law and Legal Studies	6	1.166667	3	0	1.471960
	07 Agricultural and Veterinary Sciences	4	3.500000	10	0	4.509250
	46 Information and Computing Sciences	3	0.333333	1	0	0.577350
	32 Biomedical and Clinical Sciences	3	0.000000	0	0	0.000000
	21 History and Archaeology	3	3.333333	9	0	4.932883
	19 Studies in Creative Arts and Writing	3	0.000000	0	0	0.000000
	31 Biological Sciences	2	0.000000	0	0	0.000000
	35 Commerce, Management, Tourism and Services	1	1.000000	1	1	NaN
	36 Creative Arts and Writing	1	0.000000	0	0	NaN
	37 Earth Sciences	1	1.000000	1	1	NaN

	count	mean	max	min	std
Category_1					
51 Physical Sciences	1	9.000000	9	9	NaN

Iran University of Science and Technology

```
In [ ]: df[df['University'] == "Iran University of Science and Technology"].groupby('Catego
```

Out[]:

		count	mean	max	min	std
Category_1						
	09 Engineering	1696	1.650943	507	0	15.099734
	08 Information and Computing Sciences	677	0.927622	26	0	2.034268
	03 Chemical Sciences	608	1.697368	48	0	3.045594
	01 Mathematical Sciences	484	0.756198	21	0	1.562521
	11 Medical and Health Sciences	175	2.440000	59	0	5.929238
	02 Physical Sciences	170	1.088235	13	0	1.583339
	10 Technology	120	0.783333	6	0	1.189473
	04 Earth Sciences	76	2.407895	53	0	7.941751
	06 Biological Sciences	65	2.476923	34	0	5.229447
	05 Environmental Sciences	53	1.396226	15	0	2.256100
	12 Built Environment and Design	43	1.279070	14	0	2.693251
	15 Commerce, Management, Tourism and Services	43	0.604651	3	0	0.954676
	14 Economics	41	0.853659	12	0	2.080396
	17 Psychology and Cognitive Sciences	40	3.200000	59	0	9.855885
	13 Education	19	1.263158	3	0	0.805682
	16 Studies in Human Society	10	1.600000	3	0	1.173788
	00 No Category	9	0.333333	3	0	1.000000
	20 Language, Communication and Culture	8	0.625000	3	0	1.060660
	07 Agricultural and Veterinary Sciences	4	0.000000	0	0	0.000000
	32 Biomedical and Clinical Sciences	3	0.000000	0	0	0.000000
	46 Information and Computing Sciences	3	0.666667	1	0	0.577350
	22 Philosophy and Religious Studies	2	0.500000	1	0	0.707107
	35 Commerce, Management, Tourism and Services	2	2.000000	3	1	1.414214
	21 History and Archaeology	1	3.000000	3	3	NaN
	34 Chemical Sciences	1	0.000000	0	0	NaN
	40 Engineering	1	0.000000	0	0	NaN
	47 Language, Communication and Culture	1	0.000000	0	0	NaN
	49 Mathematical Sciences	1	0.000000	0	0	NaN

University of Tabriz

In []: df[df['University'] == "University of Tabriz"].groupby('Category_1')[['Altmetric_Att

Out[]:

		count	mean	max	min	std
Category_1						
	09 Engineering	856	1.386682	78	0	3.819771
	03 Chemical Sciences	676	1.390533	16	0	1.822046
	11 Medical and Health Sciences	597	4.294807	349	0	20.396165
	06 Biological Sciences	524	1.979008	33	0	3.172772
	08 Information and Computing Sciences	286	1.146853	11	0	1.732894
	01 Mathematical Sciences	266	0.725564	8	0	1.103966
	02 Physical Sciences	265	1.769811	57	0	4.081072
	04 Earth Sciences	205	1.946341	57	0	5.777072
	07 Agricultural and Veterinary Sciences	115	3.234783	202	0	18.886703
	05 Environmental Sciences	111	1.378378	19	0	2.475531
	17 Psychology and Cognitive Sciences	51	2.764706	20	0	3.957718
	14 Economics	44	0.295455	3	0	0.667503
	10 Technology	38	2.763158	31	0	6.292141
	12 Built Environment and Design	22	1.681818	8	0	2.056033
	15 Commerce, Management, Tourism and Services	16	1.062500	4	0	1.340087
	16 Studies in Human Society	11	0.636364	3	0	1.026911
	13 Education	10	4.600000	28	1	8.369256
	40 Engineering	9	0.444444	3	0	1.013794
	00 No Category	8	0.500000	3	0	1.069045
	20 Language, Communication and Culture	7	2.142857	6	0	2.410295
	22 Philosophy and Religious Studies	4	2.250000	6	0	2.872281
	31 Biological Sciences	2	1.500000	2	1	0.707107
	37 Earth Sciences	2	0.500000	1	0	0.707107
	18 Law and Legal Studies	2	0.500000	1	0	0.707107
	49 Mathematical Sciences	1	0.000000	0	0	NaN
	51 Physical Sciences	1	0.000000	0	0	NaN
	52 Psychology	1	1.000000	1	1	NaN

Tarbiat Modares University

In []: df[df['University'] == "Tarbiat Modares University"].groupby('Category_1')['Altmetr

		count	mean	max	min	std
Category_1						
	11 Medical and Health Sciences	2728	5.725073	1367	0	39.696113
	06 Biological Sciences	1634	3.654835	554	0	21.676706
	03 Chemical Sciences	1219	1.759639	191	0	5.853618
	09 Engineering	1166	1.572899	182	0	6.109256
	08 Information and Computing Sciences	398	1.153266	13	0	1.814891
	04 Earth Sciences	331	2.540785	224	0	17.122217
	01 Mathematical Sciences	246	1.150407	14	0	1.924530
	05 Environmental Sciences	239	11.364017	961	0	74.816002
	02 Physical Sciences	194	1.376289	13	0	2.009731
	10 Technology	136	1.794118	50	0	4.960761
	07 Agricultural and Veterinary Sciences	135	1.600000	11	0	2.130798
	17 Psychology and Cognitive Sciences	86	3.430233	65	0	9.368961
	14 Economics	62	3.048387	58	0	8.136998
	16 Studies in Human Society	62	4.532258	136	0	17.614932
	15 Commerce, Management, Tourism and Services	59	1.779661	60	0	7.900388
	13 Education	56	2.000000	21	0	3.668044
	12 Built Environment and Design	32	1.968750	36	0	6.286105
	22 Philosophy and Religious Studies	17	4.117647	19	0	5.765006
	00 No Category	15	1.066667	9	0	2.685056
	20 Language, Communication and Culture	13	3.461538	31	0	8.362769
	21 History and Archaeology	12	241.166667	2043	0	610.841719
	32 Biomedical and Clinical Sciences	7	1.000000	5	0	1.825742
	40 Engineering	6	1.333333	3	0	1.366260
	31 Biological Sciences	3	1.000000	2	0	1.000000
	46 Information and Computing Sciences	3	2.333333	6	0	3.214550
	19 Studies in Creative Arts and Writing	2	2.000000	3	1	1.414214
	18 Law and Legal Studies	2	22.000000	43	1	29.698485
	34 Chemical Sciences	1	2.000000	2	2	NaN
	35 Commerce, Management, Tourism and Services	1	0.000000	0	0	NaN
	37 Earth Sciences	1	1.000000	1	1	NaN
	41 Environmental Sciences	1	1.000000	1	1	NaN
	42 Health Sciences	1	1.000000	1	1	NaN

	count	mean	max	min	std
Category_1					
44 Human Society	1	2.000000	2	2	NaN
52 Psychology	1	1.000000	1	1	NaN

University of Tehran

```
In [ ]: df[df['University'] == "University of Tehran"].groupby('Category_1')['Altmetric_Att
```

Out[]:

		count	mean	max	min	std
Category_1						
	09 Engineering	3134	1.620294	661	0	12.300163
	11 Medical and Health Sciences	2564	7.201248	4568	0	93.731110
	06 Biological Sciences	2469	6.144998	2766	0	66.144855
	03 Chemical Sciences	1736	1.908410	60	0	3.682534
	08 Information and Computing Sciences	1320	1.409848	177	0	5.738639
	04 Earth Sciences	919	6.078346	3149	0	104.804675
	01 Mathematical Sciences	847	1.573790	109	0	5.616080
	02 Physical Sciences	614	2.464169	92	0	6.164508
	05 Environmental Sciences	474	4.324895	218	0	16.078689
	07 Agricultural and Veterinary Sciences	375	1.640000	63	0	3.780190
	17 Psychology and Cognitive Sciences	314	8.219745	333	0	32.669701
	15 Commerce, Management, Tourism and Services	268	1.119403	24	0	2.681731
	16 Studies in Human Society	236	2.830508	44	0	5.868495
	10 Technology	199	1.417085	67	0	5.005241
	14 Economics	191	4.916230	409	0	31.699133
	00 No Category	98	0.602041	7	0	1.419596
	13 Education	90	3.444444	66	0	8.968247
	12 Built Environment and Design	81	1.382716	22	0	2.790913
	21 History and Archaeology	69	58.202899	2043	0	302.322120
	20 Language, Communication and Culture	56	1.928571	31	0	4.548155
	22 Philosophy and Religious Studies	34	1.352941	4	0	1.432995
	32 Biomedical and Clinical Sciences	20	0.500000	3	0	1.100239
	31 Biological Sciences	16	1.562500	5	0	1.504161
	40 Engineering	15	1.266667	10	0	2.631313
	18 Law and Legal Studies	10	12.300000	56	0	20.265186
	46 Information and Computing Sciences	9	0.333333	1	0	0.500000
	37 Earth Sciences	7	0.571429	3	0	1.133893
	19 Studies in Creative Arts and Writing	4	5.000000	10	1	3.915780
	30 Agricultural, Veterinary and Food Sciences	4	3.000000	9	0	4.082483
	44 Human Society	4	1.750000	3	1	0.957427
	33 Built Environment and Design	3	1.333333	2	1	0.577350
	35 Commerce, Management, Tourism and Services	3	1.000000	1	1	0.000000

	count	mean	max	min	std
Category_1					
50 Philosophy and Religious Studies	3	23.000000	66	0	37.269290
51 Physical Sciences	2	0.000000	0	0	0.000000
49 Mathematical Sciences	1	0.000000	0	0	NaN
39 Education	1	1.000000	1	1	NaN
42 Health Sciences	1	1.000000	1	1	NaN
41 Environmental Sciences	1	1.000000	1	1	NaN
38 Economics	1	0.000000	0	0	NaN
52 Psychology	1	1.000000	1	1	NaN