Exploring and Mining Data on Journal Publishing Patterns

Sahil Sah, Veniamin Knyazev

April 17, 2023

Introduction

This report presents the results of the following analytical queries against our MySQL data warehouse, aiming to provide insights into the publishing patterns of various journals:

1. Top five journals with the most articles published in them for the time period.
2. Number of articles per journal per year broken down by quarter.

Setup for MySQL Instance

Analytical Query I: Top five journals with the most articles published in them for the time period.

We identified the top five journals with the most articles published in them for the selected time period. Understanding the popularity and influence of these journals can help analysts and managers identify trends and make informed decisions.

```
# Store the result to a dataframe
top_5_journals <- dbGetQuery(mysqlCon,
   "SELECT jd.title AS journal_title, SUM(jf.articles_count) AS sumOfArticles
    FROM Journal_Facts jf JOIN Journal_Dim jd ON jf.journal_id = jd.journal_id
    GROUP BY jd.title
    ORDER BY sumOfArticles DESC
    LIMIT 5")
top_5_journals
```

```
##                             journal_title sumOfArticles
## 1 The Journal of pharmacy and pharmacology          1015
## 2           Biochimica et biophysica acta           904
## 3     The Journal of biological chemistry           595
## 4   Annales de l'anesthesiologie francaise           523
## 5                            Biochemistry           370
```

Analytical Query II: Number of articles per journal per year broken down by quarter. Again, think of a good way to show this. Is a table better than a visualization or is it better to use a visualization.

We analyzed the number of articles per journal per year, broken down by quarter, for the top five journals. This data provides insight into the publishing patterns of different journals, allowing analysts and managers

to track trends and plan for future publication needs.

```
# Store the result to a dataframe
journal_per_year <- dbGetQuery(mysqlCon,
   "SELECT jd.title AS journal_title, jf.publication_year, jf.publication_quarter, SUM(jf.articles_coun
    FROM Journal_Facts jf JOIN Journal_Dim jd ON jf.journal_id = jd.journal_id
    WHERE jd.title IN (
      SELECT title
      FROM (
        SELECT jd.title, SUM(jf.articles_count) AS sumOfArticles
        FROM Journal_Facts jf JOIN Journal_Dim jd ON jf.journal_id = jd.journal_id
        GROUP BY jd.title
        ORDER BY sumOfArticles DESC
        LIMIT 5
      ) AS top_journals
    )
    GROUP BY jd.title, jf.publication_year, jf.publication_quarter")

journal_per_year
```

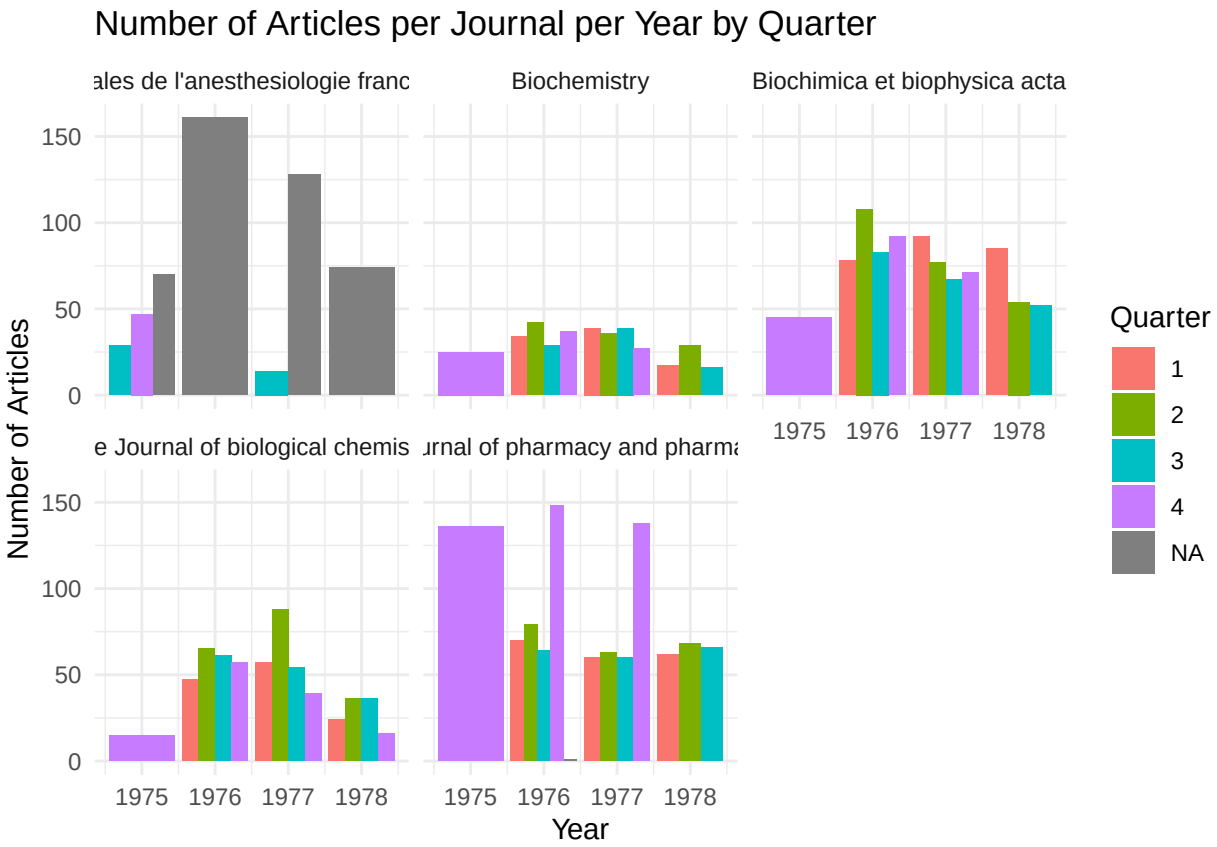
```
##                              journal_title publication_year
## 1                             Biochemistry             1975
## 2                             Biochemistry             1976
## 3                             Biochemistry             1976
## 4                             Biochemistry             1976
## 5                             Biochemistry             1976
## 6                             Biochemistry             1977
## 7                             Biochemistry             1977
## 8                             Biochemistry             1977
## 9                             Biochemistry             1977
## 10                            Biochemistry             1978
## 11                            Biochemistry             1978
## 12                            Biochemistry             1978
## 13           Biochimica et biophysica acta             1975
## 14           Biochimica et biophysica acta             1976
## 15           Biochimica et biophysica acta             1976
## 16           Biochimica et biophysica acta             1976
## 17           Biochimica et biophysica acta             1976
## 18           Biochimica et biophysica acta             1977
## 19           Biochimica et biophysica acta             1977
## 20           Biochimica et biophysica acta             1977
## 21           Biochimica et biophysica acta             1977
## 22           Biochimica et biophysica acta             1978
## 23           Biochimica et biophysica acta             1978
## 24           Biochimica et biophysica acta             1978
## 25     The Journal of biological chemistry             1975
## 26     The Journal of biological chemistry             1976
## 27     The Journal of biological chemistry             1976
## 28     The Journal of biological chemistry             1976
## 29     The Journal of biological chemistry             1976
## 30     The Journal of biological chemistry             1977
## 31     The Journal of biological chemistry             1977
## 32     The Journal of biological chemistry             1977
## 33     The Journal of biological chemistry             1977
```

```
## 34      The Journal of biological chemistry            1978
## 35      The Journal of biological chemistry            1978
## 36      The Journal of biological chemistry            1978
## 37      The Journal of biological chemistry            1978
## 38   Annales de l'anesthesiologie francaise            1975
## 39   Annales de l'anesthesiologie francaise            1975
## 40   Annales de l'anesthesiologie francaise            1975
## 41   Annales de l'anesthesiologie francaise            1976
## 42   Annales de l'anesthesiologie francaise            1977
## 43   Annales de l'anesthesiologie francaise            1977
## 44   Annales de l'anesthesiologie francaise            1978
## 45 The Journal of pharmacy and pharmacology            1975
## 46 The Journal of pharmacy and pharmacology            1976
## 47 The Journal of pharmacy and pharmacology            1976
## 48 The Journal of pharmacy and pharmacology            1976
## 49 The Journal of pharmacy and pharmacology            1976
## 50 The Journal of pharmacy and pharmacology            1976
## 51 The Journal of pharmacy and pharmacology            1977
## 52 The Journal of pharmacy and pharmacology            1977
## 53 The Journal of pharmacy and pharmacology            1977
## 54 The Journal of pharmacy and pharmacology            1977
## 55 The Journal of pharmacy and pharmacology            1978
## 56 The Journal of pharmacy and pharmacology            1978
## 57 The Journal of pharmacy and pharmacology            1978
##    publication_quarter sumPerQuarter
## 1                    4            25
## 2                    1            34
## 3                    2            42
## 4                    3            29
## 5                    4            37
## 6                    1            39
## 7                    2            36
## 8                    3            39
## 9                    4            27
## 10                   1            17
## 11                   2            29
## 12                   3            16
## 13                   4            45
## 14                   1            78
## 15                   2           108
## 16                   3            83
## 17                   4            92
## 18                   1            92
## 19                   2            77
## 20                   3            67
## 21                   4            71
## 22                   1            85
## 23                   2            54
## 24                   3            52
## 25                   4            15
## 26                   1            47
## 27                   2            65
## 28                   3            61
## 29                   4            57
```

```
## 30                    1            57
## 31                    2            88
## 32                    3            54
## 33                    4            39
## 34                    1            24
## 35                    2            36
## 36                    3            36
## 37                    4            16
## 38                   NA            70
## 39                    3            29
## 40                    4            47
## 41                   NA           161
## 42                   NA           128
## 43                    3            14
## 44                   NA            74
## 45                    4           136
## 46                   NA             1
## 47                    1            70
## 48                    2            79
## 49                    3            64
## 50                    4           148
## 51                    1            60
## 52                    2            63
## 53                    3            60
## 54                    4           138
## 55                    1            62
## 56                    2            68
## 57                    3            66
```

We used ggplot2 to create a bar chart showing the number of articles per journal per year, broken down by quarter, for the top five journals. This visualization helps users quickly understand the results of our analysis.

```
ggplot(data = journal_per_year, aes(x = publication_year, y = sumPerQuarter, fill = factor(publication_
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(~ journal_title) +
  labs(title = "Number of Articles per Journal per Year by Quarter",
       x = "Year",
       y = "Number of Articles",
       fill = "Quarter") +
  theme_minimal()
```



Conclusion

In this report, we presented the results of two analytical queries on our MySQL data warehouse. We identified the top five journals with the most articles published in them for the selected time period and analyzed the number of articles per journal per year, broken down by quarter. By presenting the data in both tabular and visual formats, we have made it easy for users to quickly understand and interpret the results of our analysis.

These insights can help analysts and managers track trends, understand the popularity and influence of specific journals, and plan for future publication needs.