```
In [1]:  # inputting libraries
         import numpy as np
         import pandas as pd
         import matplotlib.pyplot as plt
         import seaborn as sns
         %matplotlib inline
         sns.set(color_codes=True)
```

```
In [2]:  kaggle = pd.read_excel("E:/R-programming/Kaggle_Books.xlsx")
         kaggle
```

Out[2]:

	ID	Purchase_Date	Time	Quater	Year	Month	Special_Day	Online_Sale_Offers	Customer
0	1	2018-01-01	08:57:00	1	2018	1	1	1	B07FWXV
1	2	2018-01-02	02:04:00	1	2018	1	1	1	B07FWVX
2	3	2018-01-03	05:08:00	1	2018	1	1	1	B07FWV
3	4	2018-01-04	16:06:00	1	2018	1	0	1	B07FWYX
4	5	2018-01-04	20:12:00	1	2018	1	0	0	B07FWXA
...	...	...	...	...	...	...	...	...	
3475	3476	2019-12-31	12:02:00	4	2019	12	1	1	B07FWV
3476	3477	2019-12-31	02:53:00	4	2019	12	1	1	B07FWV
3477	3478	2019-12-31	06:02:00	4	2019	12	1	1	B07FWYX
3478	3479	2019-12-31	07:28:00	4	2019	12	1	1	B07FWVV
3479	3480	2019-12-31	09:06:00	4	2019	12	1	1	B07FWWS

3480 rows × 26 columns

```
In [3]:  kaggle.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3480 entries, 0 to 3479
Data columns (total 26 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   ID                  3480 non-null   int64
 1   Purchase_Date       3480 non-null   datetime64[ns]
 2   Time                3480 non-null   object
 3   Quater              3480 non-null   int64
 4   Year                3480 non-null   int64
 5   Month               3480 non-null   int64
 6   Special_Day         3480 non-null   int64
 7   Online_Sale_Offers  3480 non-null   int64
 8   Customer_ID         3480 non-null   object
 9   Gender              3480 non-null   object
 10  Product_Name        3480 non-null   object
 11  Item_Status         3480 non-null   object
 12  Quantity            3480 non-null   int64
 13  Currency            3480 non-null   object
 14  Item_Price          3480 non-null   float64
 15  Shipping_Price      3480 non-null   float64
 16  Ship_City           3480 non-null   object
 17  Ship_State          3480 non-null   object
 18  Ship_Postal_Code    3480 non-null   int64
 19  Category            3480 non-null   object
 20  Total_amount        3480 non-null   float64
 21  Author              3480 non-null   object
 22  Publication         3480 non-null   object
 23  Profit_Percentage(%) 3480 non-null  int64
 24  Profit(INR)         3480 non-null   float64
 25  Cost_Price          3480 non-null   float64
dtypes: datetime64[ns](1), float64(5), int64(9), object(11)
memory usage: 707.0+ KB
```

In [4]:
```python
kaggle.isnull().sum()
```

Out[4]:
```
ID                          0
Purchase_Date               0
Time                        0
Quater                      0
Year                        0
Month                       0
Special_Day                 0
Online_Sale_Offers          0
Customer_ID                 0
Gender                      0
Product_Name                0
Item_Status                 0
Quantity                    0
Currency                    0
Item_Price                  0
Shipping_Price              0
Ship_City                   0
Ship_State                  0
Ship_Postal_Code            0
Category                    0
Total_amount                0
Author                      0
Publication                 0
Profit_Percentage(%)        0
Profit(INR)                 0
Cost_Price                  0
dtype: int64
```

In [5]: `kaggle.dtypes`

Out[5]:
```
ID                          int64
Purchase_Date               datetime64[ns]
Time                        object
Quater                      int64
Year                        int64
Month                       int64
Special_Day                 int64
Online_Sale_Offers          int64
Customer_ID                 object
Gender                      object
Product_Name                object
Item_Status                 object
Quantity                    int64
Currency                    object
Item_Price                  float64
Shipping_Price              float64
Ship_City                   object
Ship_State                  object
Ship_Postal_Code            int64
Category                    object
Total_amount                float64
Author                      object
Publication                 object
Profit_Percentage(%)        int64
Profit(INR)                 float64
Cost_Price                  float64
dtype: object
```

In [6]:
```python
#quantity by purchase date
top_date= kaggle.groupby('Purchase_Date')['Quantity'].sum().sort_values(ascending = Fa
```

```
top_date
```

Out[6]:
```
Purchase_Date
2019-04-07    55
2019-06-01    47
2019-04-03    47
2019-04-04    35
2019-07-27    35
2019-05-16    33
2019-08-09    32
2019-05-15    31
2019-07-10    29
2019-07-24    25
Name: Quantity, dtype: int64
```

In [7]:
```python
top_date.plot(kind = 'line', title = 'TOP DATES BY QUANTITY OF BOOKS PURCHASED', ylabe
```

Out[7]:
```
<Axes: title={'center': 'TOP DATES BY QUANTITY OF BOOKS PURCHASED'}, xlabel='Purchase
_Date', ylabel='quantity purchased'>
```


In [8]:
```python
#quantity by quater
top_quater= kaggle.groupby('Quater')['Quantity'].sum().sort_values(ascending = False)
top_quater
```

Out[8]:
```
Quater
2    1574
3    1080
1     589
4     412
Name: Quantity, dtype: int64
```

In [9]:
```python
top_quater.plot(kind = 'bar', title = 'QUATERS BY QUANTITY OF BOOKS PURCHASED',ylabel
```

Out[9]: `<Axes: title={'center': 'QUATERS BY QUANTITY OF BOOKS PURCHASED'}, xlabel='Quater', y label='total books purchased'>`

QUATERS BY QUANTITY OF BOOKS PURCHASED

In [10]:
```python
#quantity by year
top_year = kaggle.groupby('Year')['Quantity'].sum().sort_values(ascending = False)
top_year
```

Out[10]:
```
Year
2019    2408
2018    1247
Name: Quantity, dtype: int64
```

In [11]: `top_year.plot(kind = 'bar', title = 'QUANTITY OF BOOKS PURCHASED BY YEAR',ylabel = 'tc`

Out[11]: `<Axes: title={'center': 'QUANTITY OF BOOKS PURCHASED BY YEAR'}, xlabel='Year', ylabel ='total books purchased'>`

QUANTITY OF BOOKS PURCHASED BY YEAR

In [12]:
```python
#quantity by month
top_month= kaggle.groupby('Month')['Quantity'].sum().sort_values(ascending = False)
top_month
```

Out[12]:
```
Month
4     608
5     561
7     545
6     405
8     304
3     257
9     231
1     170
2     162
11    148
12    133
10    131
Name: Quantity, dtype: int64
```

In [13]:
```python
top_month.plot(kind = 'bar', title = 'QUANTITY OF BOOKS PURCHASED BY MONTH', ylabel ='
```

Out[13]:
```
<Axes: title={'center': 'QUANTITY OF BOOKS PURCHASED BY MONTH'}, xlabel='Month', ylab
el='total books purchased'>
```

QUANTITY OF BOOKS PURCHASED BY MONTH
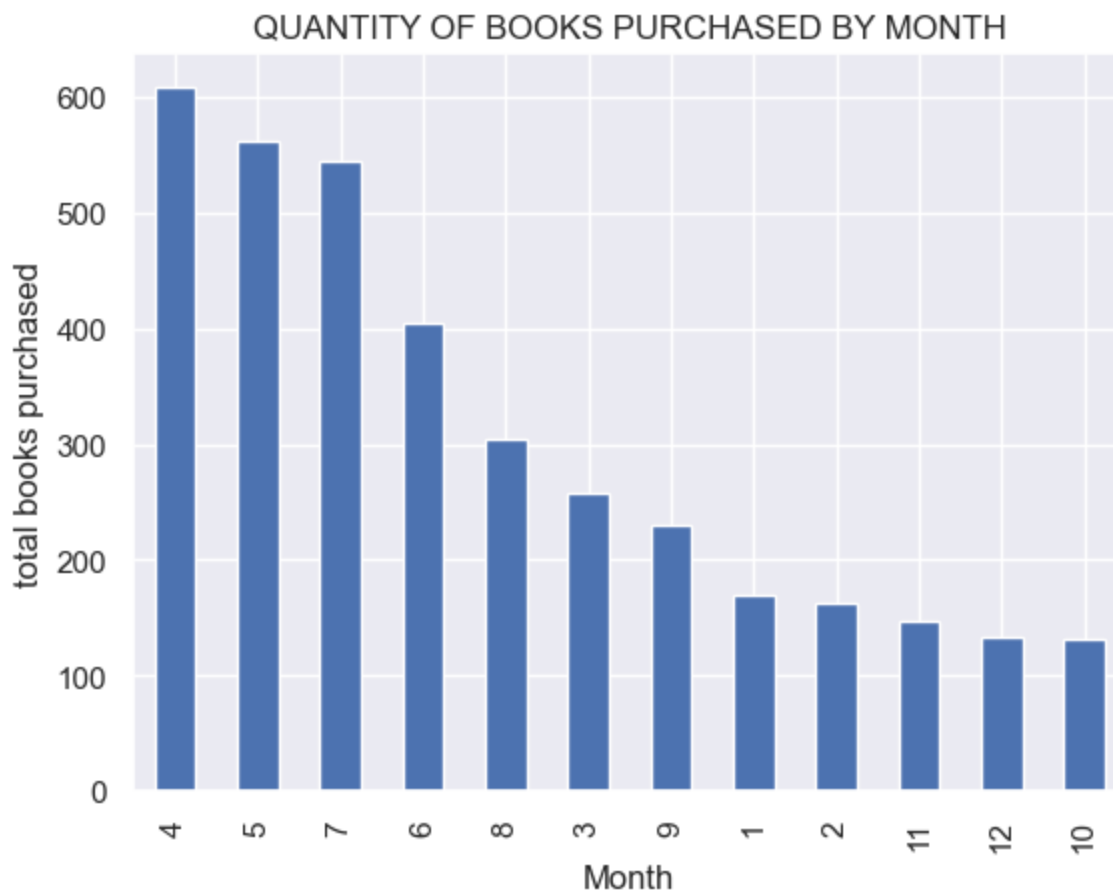

```
In [14]: #quantity by gender
         top_gender= kaggle.groupby('Gender')['Quantity'].sum().sort_values(ascending = False)
         top_gender
```

```
Out[14]: Gender
         M    1908
         F    1747
         Name: Quantity, dtype: int64
```

```
In [15]: top_gender.plot(kind = 'bar', title = 'QUANTITY OF BOOKS PURCHASED BY GENDER', ylabel
```

```
Out[15]: <Axes: title={'center': 'QUANTITY OF BOOKS PURCHASED BY GENDER'}, xlabel='Gender', yl
         abel='total books purchased'>
```

QUANTITY OF BOOKS PURCHASED BY GENDER

In [16]:
```python
#quantity by product(book) name
top_product= kaggle.groupby('Product_Name')['Quantity'].sum().sort_values(ascending =
top_product
```

Out[16]:
```
Product_Name
Lakshya Rajasthan Ki Kala, Sanskriti Va Itihas 2019 Bhag-2              246
Science Textbook for Class - 6  - 652                                   209
Science textbook for Class -7 - 2020 (NCERT)                           166
Bal RamKatha - TextBook in Hindi for Class - 6  - 645                  157
Chemistry Textbook Part - 2 for Class - 12 with Free Car Anti Slip Mat 152
Mathematics Textbook for Class - 6  - 650                              131
Words And Expressions 1 For Class 9- 976                                92
English 9th class workbook Words & Expression Ncert                     79
Ruchira Bhag - 1 Sanskrit Textbook for Class - 6  - 649                 78
Accountancy Financial Accounting Part - 1 Textbook for Class - 11- 11110 77
Name: Quantity, dtype: int64
```

In [17]:
```python
top_product.plot(kind = 'bar', title = 'QUANTITY OF TOP 10 BOOKS PURCHASED BY PRODUCT
```

Out[17]:
```
<Axes: title={'center': 'QUANTITY OF TOP 10 BOOKS PURCHASED BY PRODUCT NAME'}, xlabel
='Product_Name', ylabel='total books purchased'>
```

QUANTITY OF TOP 10 BOOKS PURCHASED BY PRODUCT NAME
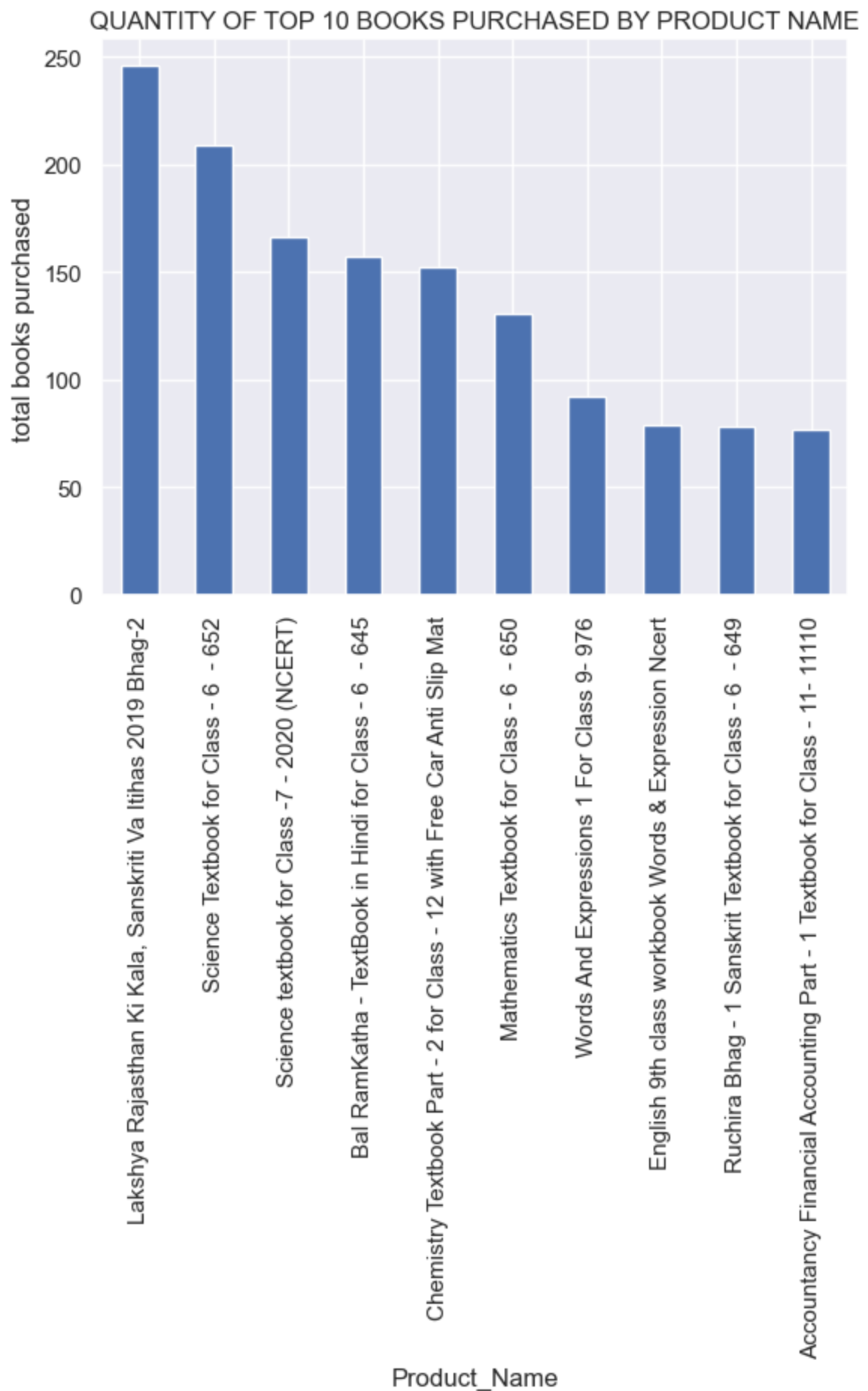
In [18]:
```python
#quantity by product
top_product= kaggle.groupby('Product_Name')['Quantity'].sum().sort_values(ascending =
top_product
```

Out[18]:
```
Product_Name
Guide to RRB Electrical Engg. (Senior Section Engineer) 2016
1
Data Structures and Algorithms Made Easy : Second Edition: Data Structure and Algorit
hmic Puzzles                                                      1
Trueman's Elementary Biology for class 11 and NEET - Vol. I (2020 Edition)
1
Kritika Bhag - 2 TextBook in Hindi for Class 10  - 1056
1
Collins Easy Learner's Dictionary
1
Lakshay PTET Pre B.ED
1
T.S. Grewal's Double Entry Book Keeping  - CBSE XII (Vol. 1: Accounting for Partnersh
ip Firms): Textbook for CBSE Class XII (2019-19 Session)     1
Theory Practice Animal Taxonomy Bio 8e
1
Classical Electrodynamics, 3ed
1
Civil Engineering (O.T.) (Objective Type)
1
Chronicle Year Book 2020 (English)
1
English Course Communicative: Main Course Book Interact in English - Class 9
1
Macro Economic Theory
1
Macroeconomics
1
Business Environment
1
T.S. Grewal's Double Entry Book Keeping  - CBSE XI (Financial Accounting): Textbook f
or CBSE Class XI                                                  1
Building Construction
1
Informatics Practices: ATextbook for Class XII
1
India a Comprehensive Geography
1
Fundamentals of Applied Statistics
1
Name: Quantity, dtype: int64
```

In [19]:
```python
#quantity by ship city
top_city= kaggle.groupby('Ship_City')['Quantity'].sum().sort_values(ascending = False)
top_city
```

Out[19]:
```
Ship_City
NEW DELHI     218
BENGALURU     191
MUMBAI        131
PUNE          105
JAIPUR         85
AHMEDABAD      74
CHENNAI        68
HYDERABAD      68
GURUGRAM       63
LUCKNOW        57
Name: Quantity, dtype: int64
```

In [20]:
```python
top_city.plot(kind = 'bar', title = 'QUANTITY OF BOOKS PURCHASED BY SHIP CITY', ylabel
```

Out[20]: <Axes: title={'center': 'QUANTITY OF BOOKS PURCHASED BY SHIP CITY'}, xlabel='Ship_Cit
y', ylabel='total books purchased'>
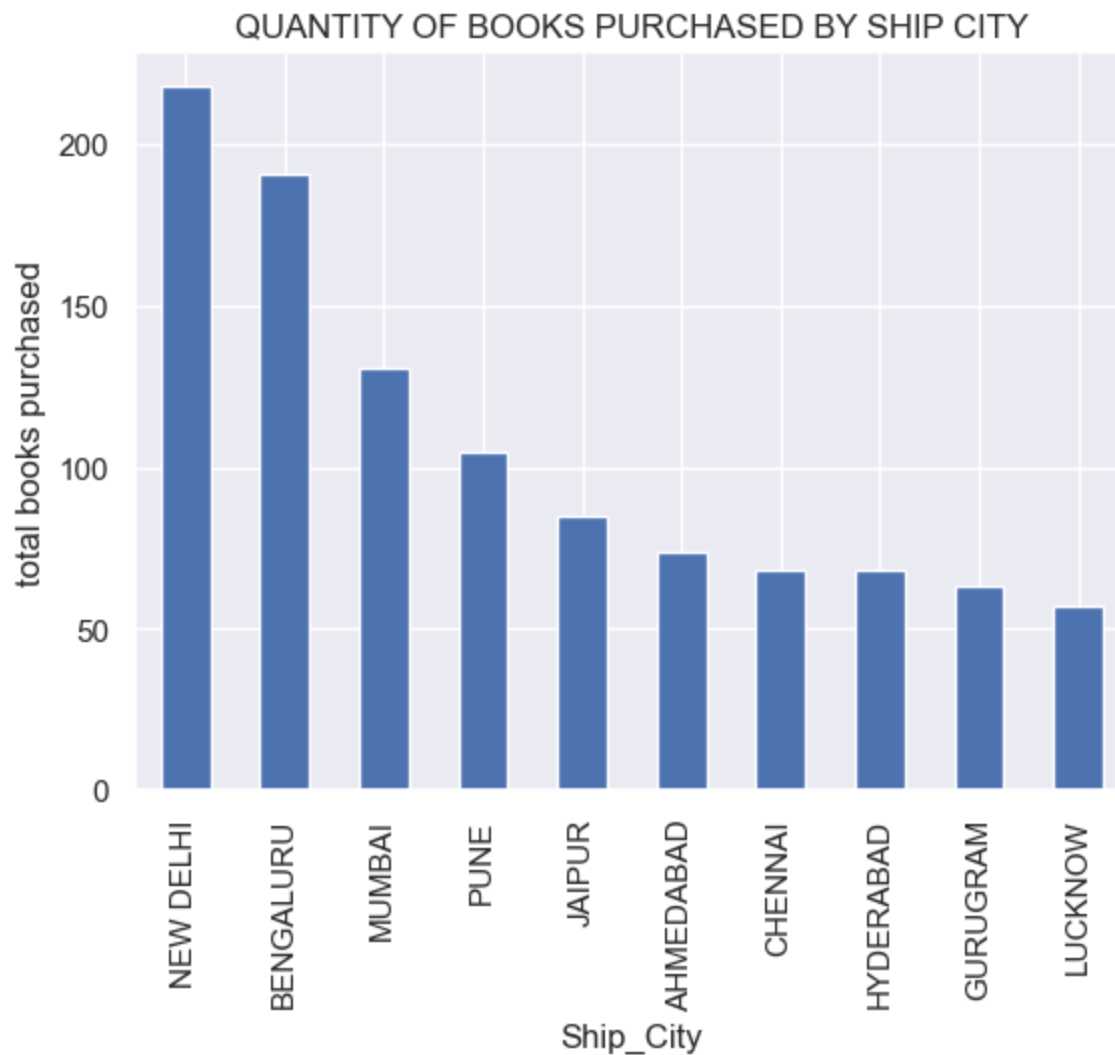
In [21]:
```python
#quantity by ship state
top_state= kaggle.groupby('Ship_State')['Quantity'].sum().sort_values(ascending = Fals
top_state
```

```
Out[21]:    Ship_State
            MAHARASHTRA       600
            UTTAR PRADESH     352
            RAJASTHAN         311
            KARNATAKA         307
            GUJARAT           270
            DELHI             259
            TAMIL NADU        216
            HARYANA           202
            WEST BENGAL       146
            TELANGANA         105
            Name: Quantity, dtype: int64
```

In [22]:   `top_state.plot(kind = 'bar', title = 'QUANTITY OF BOOKS PURCHASED BY SHIP STATE', ylat`

Out[22]:   `<Axes: title={'center': 'QUANTITY OF BOOKS PURCHASED BY SHIP STATE'}, xlabel='Ship_St ate', ylabel='total books purchased'>`

In [23]:
```
#quantity by category
top_category= kaggle.groupby('Category')['Quantity'].sum().sort_values(ascending = Fal
top_category
```

Out[23]:
```
Category
school          1350
competition      872
college          793
School           640
Name: Quantity, dtype: int64
```

In [24]: `top_category.plot(kind = 'bar', title = 'QUANTITY OF BOOKS PURCHASED BY CATEGORY', yla`

Out[24]: `<Axes: title={'center': 'QUANTITY OF BOOKS PURCHASED BY CATEGORY'}, xlabel='Category', ylabel='total books purchased'>`


In [25]:
```
#quantity by Author
top_author= kaggle.groupby('Author')['Quantity'].sum().sort_values(ascending = False).
top_author
```

Out[25]:
```
Author
NCERT                                       1708
MAHAVEER JAIN,ANSHUL JAIN KANTI JAIN         246
preeti agarwal, r.s. sharma, vinod gupta      79
Multiple Authors                              79
Preeti Agarwal                                76
Kanti Jain                                    72
N. Gregory Mankiw                             66
Andrew S. Tanenbaum                           60
Â Kanti Jain                                  58
R.S Aggarwal                                  49
Name: Quantity, dtype: int64
```

In [26]: `top_author.plot(kind = 'bar', title = 'QUANTITY OF BOOKS PURCHASED BY AUTHOR', ylabel`

Out[26]: `<Axes: title={'center': 'QUANTITY OF BOOKS PURCHASED BY AUTHOR'}, xlabel='Author', ylabel='total books purchased'>`
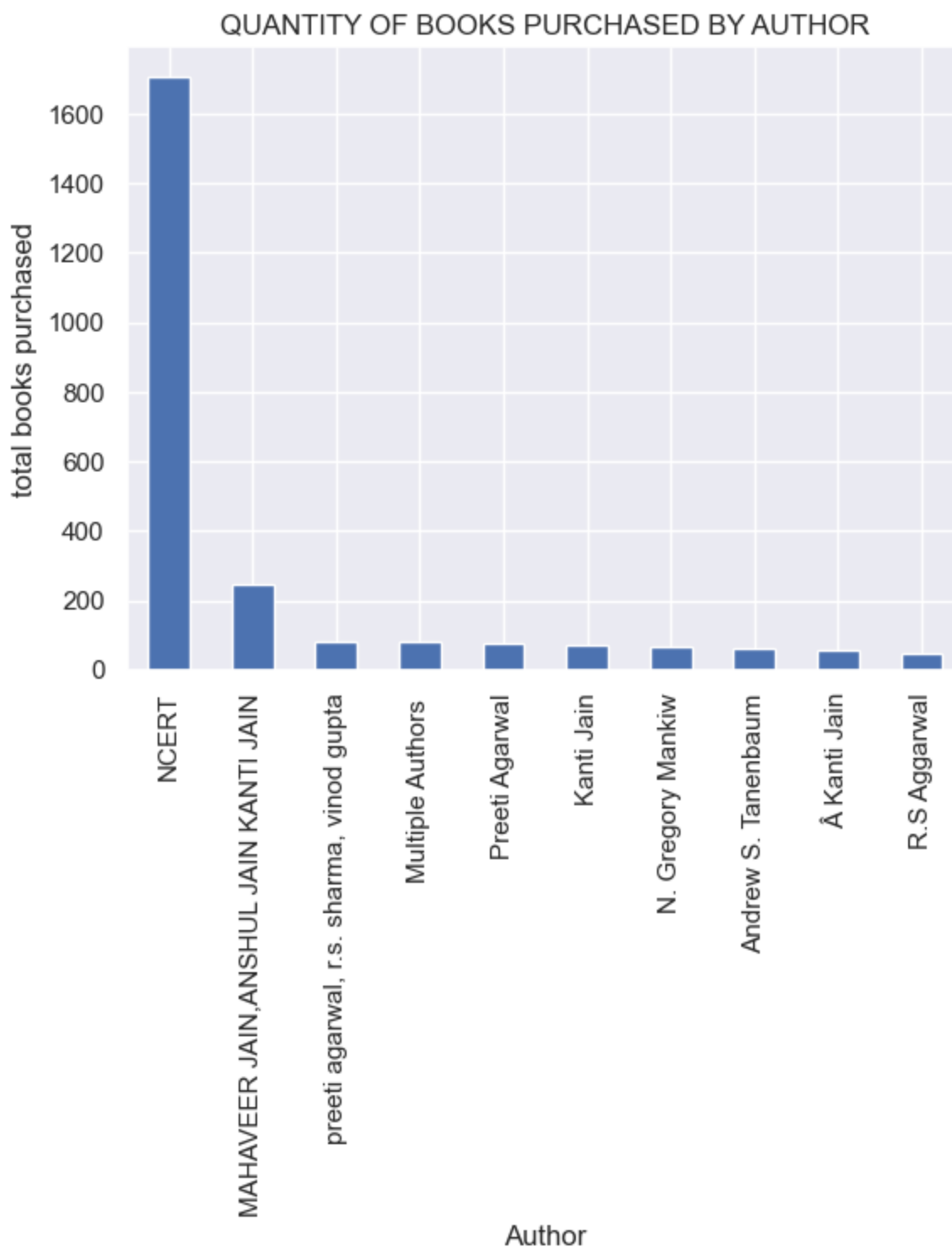

In [27]:
```python
#quantity by publication
top_publication= kaggle.groupby('Publication')['Quantity'].sum().sort_values(ascending
top_publication
```

Out[27]:
```
Publication
National Council of Education Research and Training          1421
VK Global Publications Private Limited                        255
National Council of Education Research and Training (2014)    237
Manu Prakashan Private Limited                               128
Arihant Publications                                         87
Multiple Authors Publishing                                  79
Vardhan                                                      79
Vardhan Publishers & Distributors                           76
Aravali Publication                                         66
OpenStax                                                    66
Name: Quantity, dtype: int64
```

In [28]:
```python
top_publication.plot(kind = 'bar', title = 'QUANTITY OF BOOKS PURCHASED BY PUBLICATION
```

Out[28]:
```
<Axes: title={'center': 'QUANTITY OF BOOKS PURCHASED BY PUBLICATION'}, xlabel='Public
ation', ylabel='total books purchased'>
```

QUANTITY OF BOOKS PURCHASED BY PUBLICATION
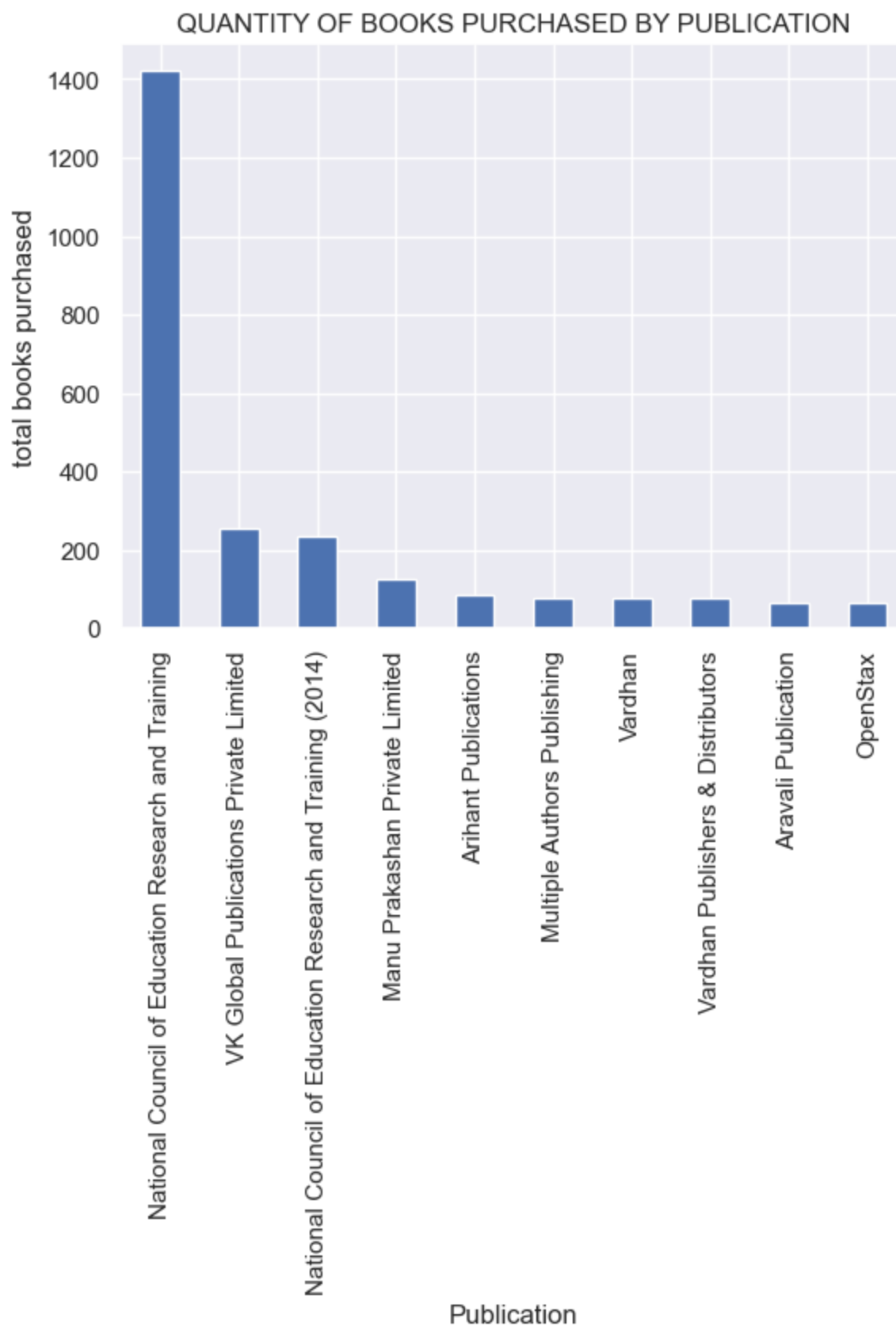


```
In [29]:  #total quantity purchased by customer
          top_customer = kaggle.groupby('Customer_ID')['Quantity'].sum().sort_values(ascending =
          top_customer
```

```
Out[29]:  Customer_ID
          B07FWWBYQ    10
          B07FWYUYQ     9
          B07FWVUVY     9
          B07FWVUSV     8
          B07FWVUUU     8
          B07FWVVWQ     8
          B07FWVWQY     7
          B07FWVVZT     7
          B07FWVYVQ     7
          B07FWVUVZ     7
          Name: Quantity, dtype: int64
```

```
In [30]:  top_customer.plot(kind = 'bar', title = 'TOP CUSTOMERS BY QUANTITY PURCHASED', ylabel
```

```
Out[30]:  <Axes: title={'center': 'TOP CUSTOMERS BY QUANTITY PURCHASED'}, xlabel='Customer_ID',
          ylabel='total quantity purchased'>
```
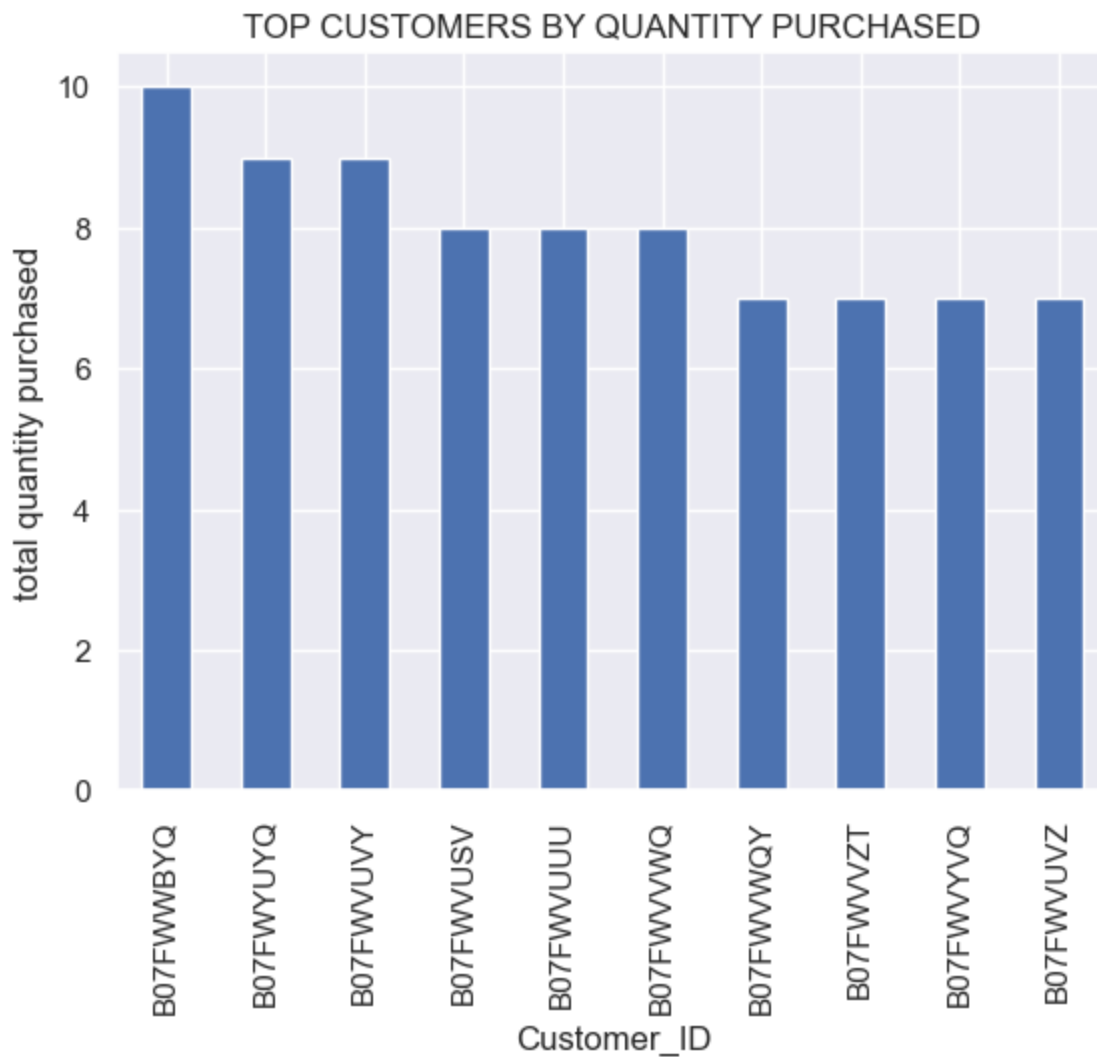


```
In [31]:  #total amount by purchase date
          top_date1= kaggle.groupby('Purchase_Date')['Total_amount'].sum().sort_values(ascending
          top_date1
```

Out[31]:
```
Purchase_Date
2019-07-28     12462.00
2018-04-09     12093.00
2019-06-01     12036.00
2019-07-27     11225.80
2018-01-15     11118.00
2019-06-07     10681.45
2018-05-13     10585.00
2018-03-07     10264.00
2019-04-07     10150.24
2018-04-07      9765.00
Name: Total_amount, dtype: float64
```

In [32]:
```python
top_date1.plot(kind = 'line', title = 'TOP DATES BY TOTAL AMOUNT SPENT ON BOOKS PURCHA
```

Out[32]:
```
<Axes: title={'center': 'TOP DATES BY TOTAL AMOUNT SPENT ON BOOKS PURCHASED'}, xlabel
='Purchase_Date', ylabel='total amount spent'>
```
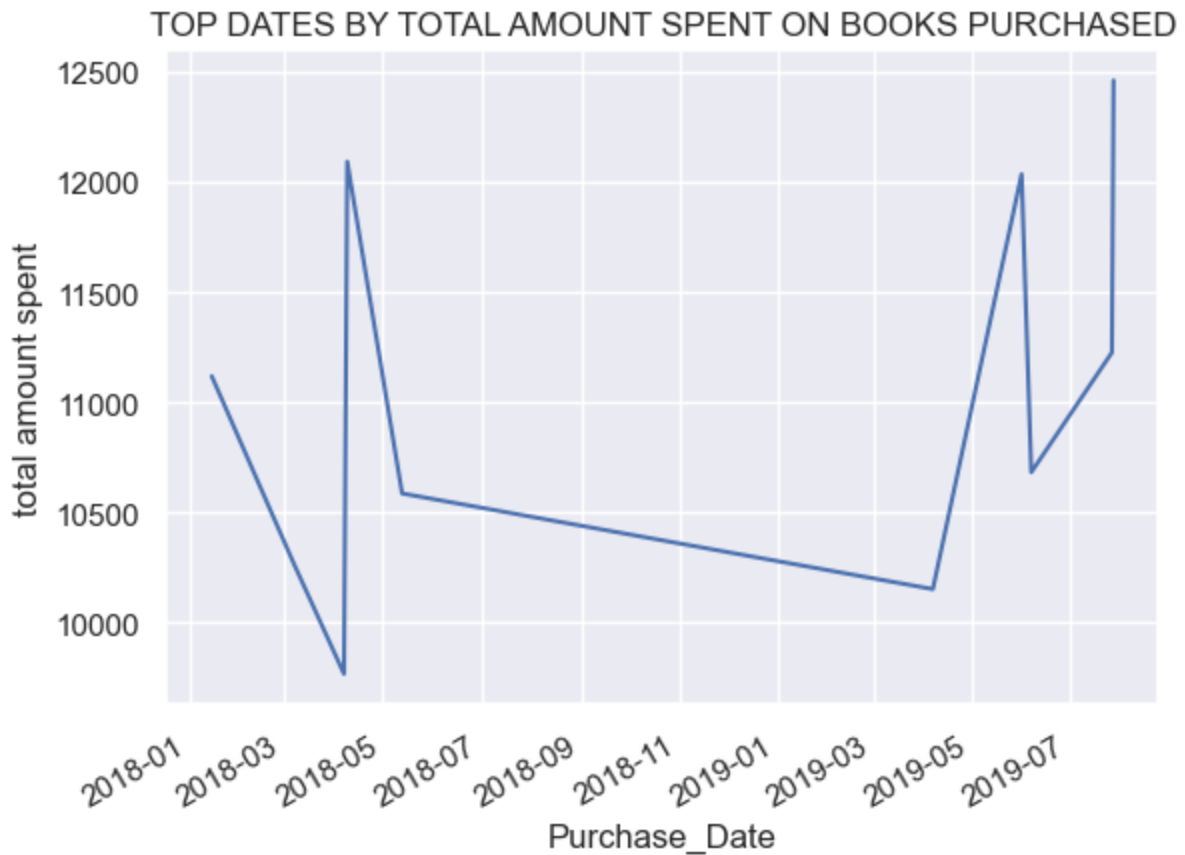

In [33]:
```python
#total amount by year
top_year1 = kaggle.groupby('Year')['Total_amount'].sum().sort_values(ascending = False
top_year1
```

Out[33]:
```
Year
2019    839793.79
2018    749789.72
Name: Total_amount, dtype: float64
```

In [34]:
```python
top_year1.plot(kind = 'bar', title = 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY YEAR',y
```

Out[34]:
```
<Axes: title={'center': 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY YEAR'}, xlabel='Yea
r', ylabel='total spent on books purchased'>
```

TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY YEAR
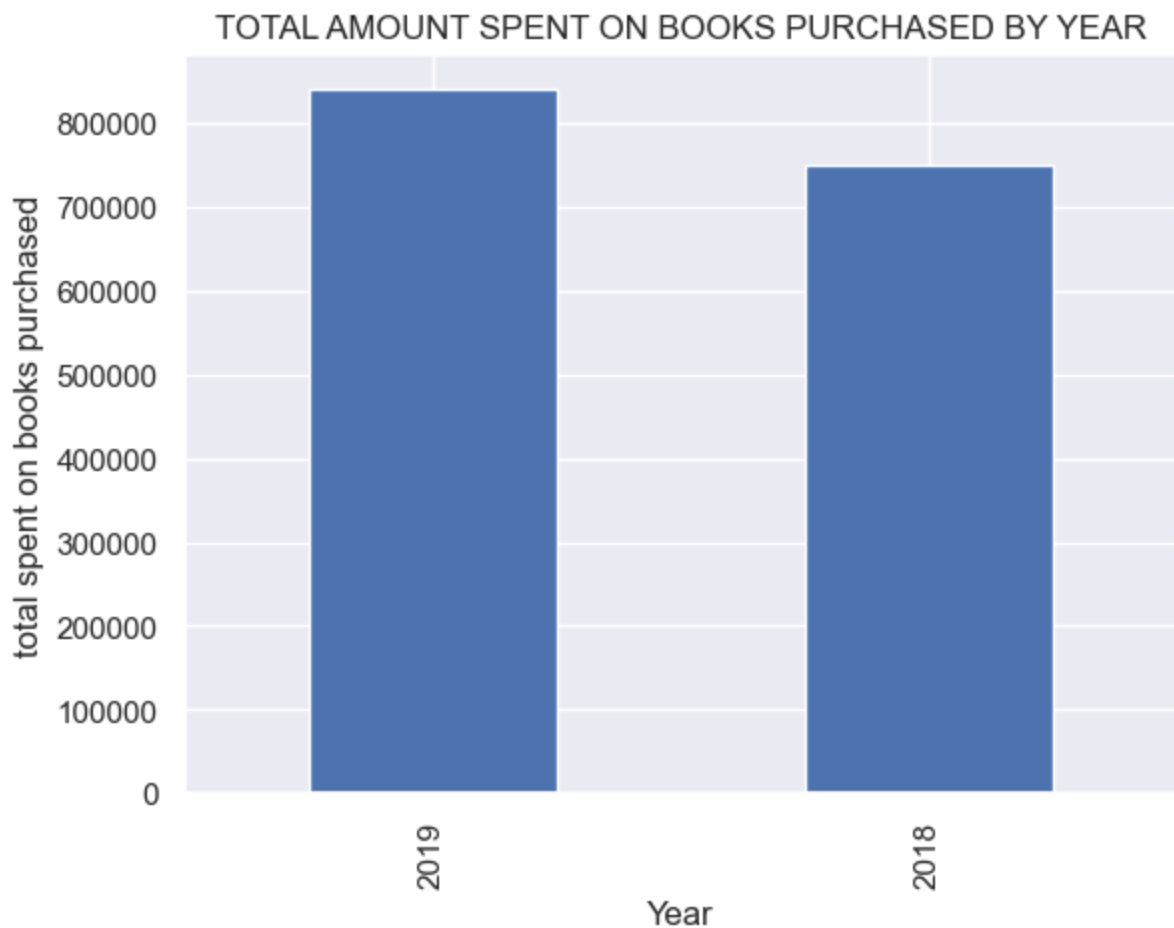

```
In [35]:   #Total amount by quater
           top_quater1= kaggle.groupby('Quater')['Total_amount'].sum().sort_values(ascending = Fa
           top_quater1
```

```
Out[35]:   Quater
           2    564338.23
           3    419632.43
           1    388978.95
           4    216633.90
           Name: Total_amount, dtype: float64
```

```
In [36]:   top_quater1.plot(kind = 'bar', title = 'QUATERS BY TOTAL AMOUNT SPENT ON BOOKS PURCHAS
```

```
Out[36]:   <Axes: title={'center': 'QUATERS BY TOTAL AMOUNT SPENT ON BOOKS PURCHASED'}, xlabel
           ='Quater', ylabel='total amount spent on books purchased'>
```

QUATERS BY TOTAL AMOUNT SPENT ON BOOKS PURCHASED
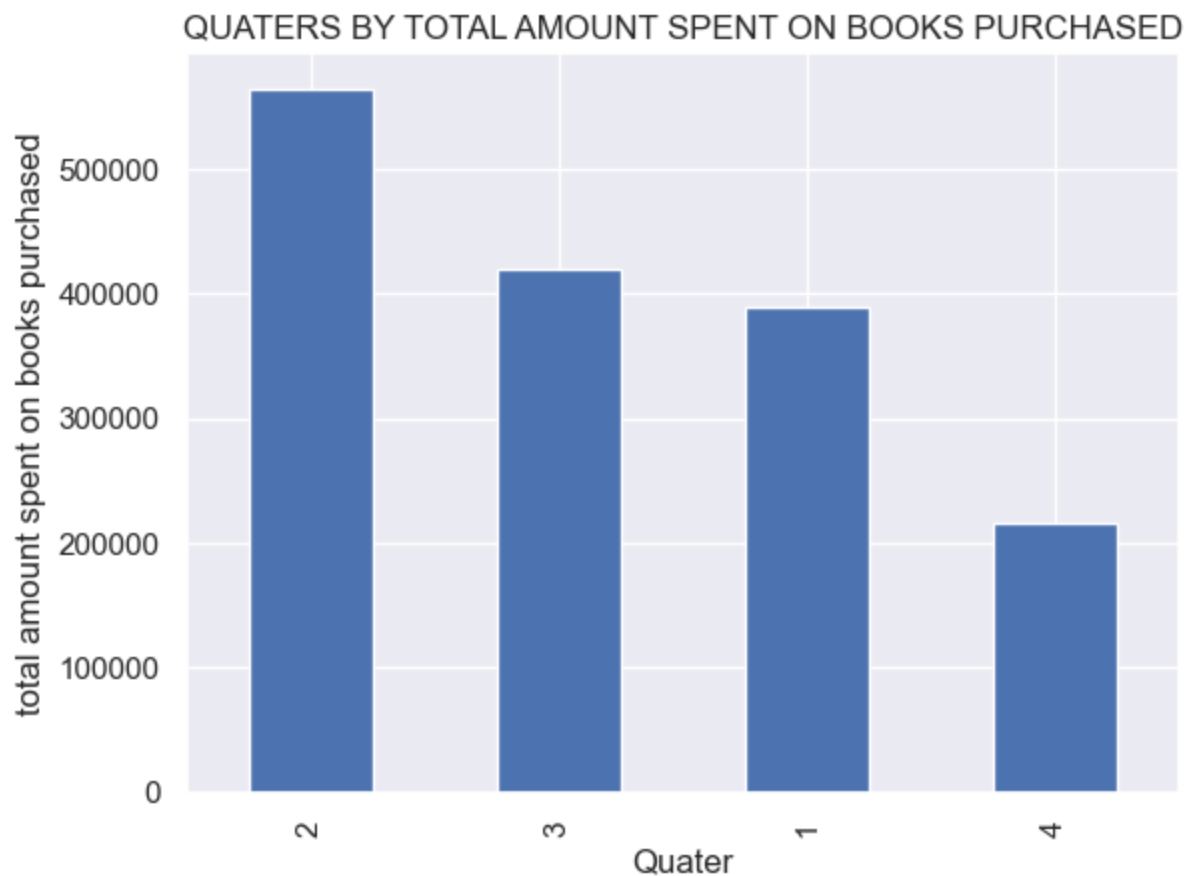

```
In [37]:  #Total_amount spent on books by month
          top_month1= kaggle.groupby('Month')['Total_amount'].sum().sort_values(ascending = Fals
          top_month1
```

```
Out[37]:  Month
          4     215493.08
          5     201417.35
          7     178028.37
          3     170001.95
          6     147427.80
          8     141504.91
          1     114498.00
          2     104479.00
          9     100099.15
          12     73566.50
          11     72277.00
          10     70790.40
          Name: Total_amount, dtype: float64
```

```
In [38]:  top_month1.plot(kind = 'bar', title = 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY MONTH'
```

```
Out[38]:  <Axes: title={'center': 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY MONTH'}, xlabel='Mo
          nth', ylabel='total amount spent on books purchased'>
```

TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY MONTH


```
In [39]:   #Total_amount spent by gender
           top_gender1= kaggle.groupby('Gender')['Total_amount'].sum().sort_values(ascending = Fa
           top_gender1
```

```
Out[39]:   Gender
           M    835050.24
           F    754533.27
           Name: Total_amount, dtype: float64
```

```
In [40]:   top_gender1.plot(kind = 'bar', title = 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY GENDE
```

```
Out[40]:   <Axes: title={'center': 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY GENDER'}, xlabel='G
           ender', ylabel='total amount spent on books purchased'>
```
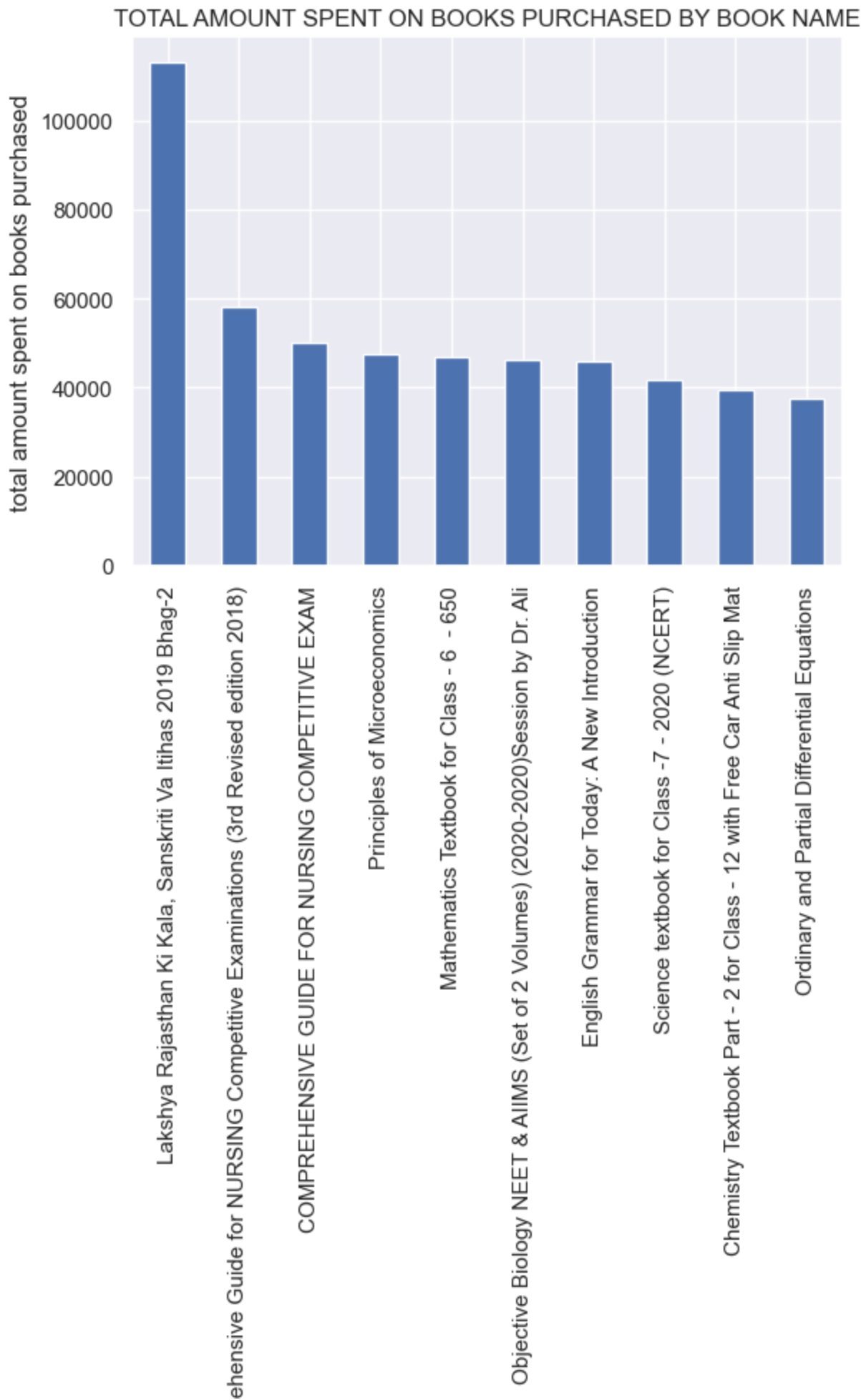
TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY GENDER


```
In [41]:   #total_amount spent by product(book) name
           top_product1= kaggle.groupby('Product_Name')['Total_amount'].sum().sort_values(ascendi
           top_product1
```

```
Out[41]:   Product_Name
           Lakshya Rajasthan Ki Kala, Sanskriti Va Itihas 2019 Bhag-2
           113102.61
           Vardhan Comprehensive Guide for NURSING Competitive Examinations (3rd Revised edition
           2018)        58307.00
           COMPREHENSIVE GUIDE FOR NURSING COMPETITIVE EXAM
           50210.00
           Principles of Microeconomics
           47661.00
           Mathematics Textbook for Class - 6  - 650
           46933.97
           Objective Biology NEET & AIIMS (Set of 2 Volumes) (2020-2020)Session by Dr. Ali
           46215.00
           English Grammar for Today: A New Introduction
           45823.00
           Science textbook for Class -7 - 2020 (NCERT)
           41641.18
           Chemistry Textbook Part - 2 for Class - 12 with Free Car Anti Slip Mat
           39658.00
           Ordinary and Partial Differential Equations
           37681.00
           Name: Total_amount, dtype: float64
```

```
In [42]:   top_product1.plot(kind = 'bar', title = 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY BOOK
```

```
Out[42]:   <Axes: title={'center': 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY BOOK NAME'}, xlabel
           ='Product_Name', ylabel='total amount spent on books purchased'>
```

TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY BOOK NAME

total amount spent on books purchased

- Lakshya Rajasthan Ki Kala, Sanskriti Va Itihas 2019 Bhag-2
- ehensive Guide for NURSING Competitive Examinations (3rd Revised edition 2018)
- COMPREHENSIVE GUIDE FOR NURSING COMPETITIVE EXAM
- Principles of Microeconomics
- Mathematics Textbook for Class - 6 - 650
- Objective Biology NEET & AIIMS (Set of 2 Volumes) (2020-2020)Session by Dr. Ali
- English Grammar for Today: A New Introduction
- Science textbook for Class -7 - 2020 (NCERT)
- Chemistry Textbook Part - 2 for Class - 12 with Free Car Anti Slip Mat
- Ordinary and Partial Differential Equations

Vardhan Compr

Product_Name

In [43]: `#Total amount spent by ship city`
`top_city1= kaggle.groupby('Ship_City')['Total_amount'].sum().sort_values(ascending = F`
`top_city1`

Out[43]:
```
Ship_City
NEW DELHI     100767.31
BENGALURU      69214.91
MUMBAI         58909.92
PUNE           47722.79
JAIPUR         45788.79
AHMEDABAD      34511.35
CHENNAI        25960.67
GURUGRAM       25624.18
HYDERABAD      24237.74
LUCKNOW        19435.24
Name: Total_amount, dtype: float64
```

In [44]: `top_city1.plot(kind = 'bar', title = 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY CITY',`

Out[44]:
```
<Axes: title={'center': 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY CITY'}, xlabel='Shi
p_City', ylabel='total amount spent on books purchased'>
```
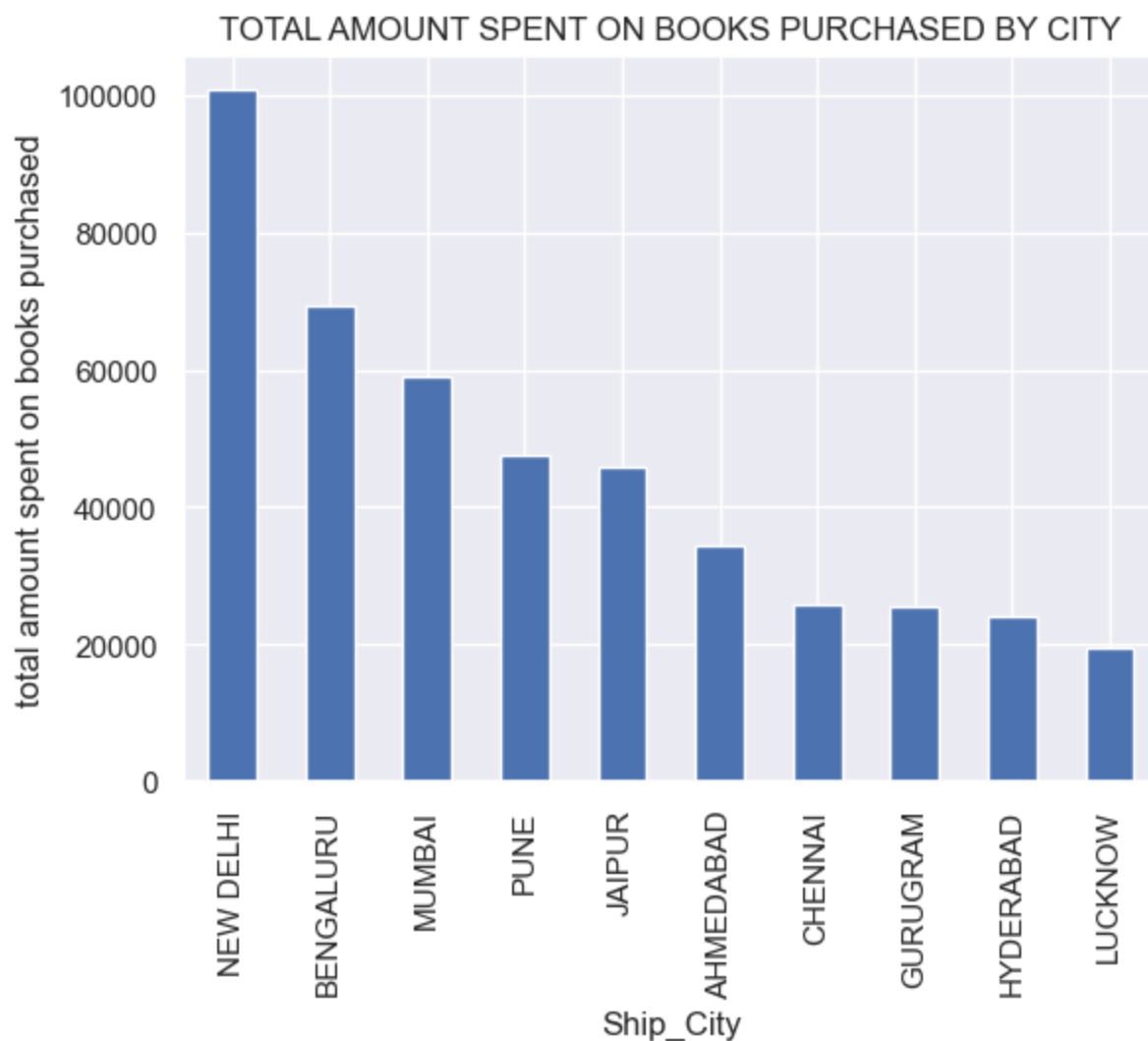
```
In [45]:  #quantity by ship state
          top_state1= kaggle.groupby('Ship_State')['Total_amount'].sum().sort_values(ascending =
          top_state1
```

```
Out[45]:  Ship_State
          MAHARASHTRA      278712.30
          RAJASTHAN        158869.23
          UTTAR PRADESH    144164.28
          DELHI            116767.26
          GUJARAT          110448.59
          KARNATAKA        105289.17
          TAMIL NADU        87567.96
          HARYANA           83314.12
          WEST BENGAL       62171.01
          TELANGANA         39947.65
          Name: Total_amount, dtype: float64
```

```
In [46]:  top_state1.plot(kind = 'bar', title = 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY STATE'
```

```
Out[46]:  <Axes: title={'center': 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY STATE'}, xlabel='Sh
          ip_State', ylabel='total amount spent on books purchased'>
```

TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY STATE
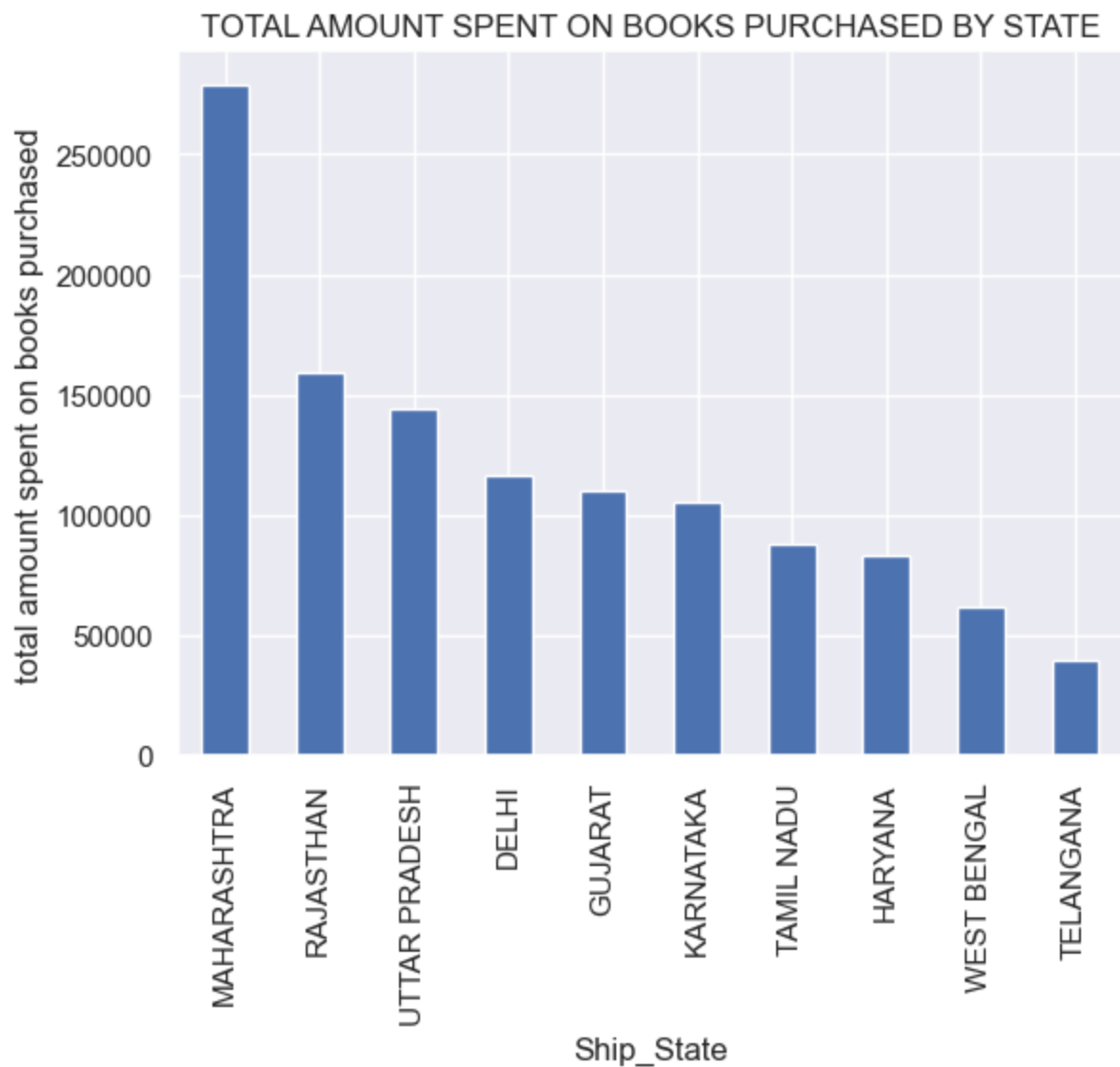

```
In [47]:  #total amount by category
          top_category1= kaggle.groupby('Category')['Total_amount'].sum().sort_values(ascending
          top_category1

Out[47]:  Category
          college        548991.50
          competition    503906.39
          school         330134.64
          School         206550.98
          Name: Total_amount, dtype: float64

In [48]:  top_category1.plot(kind = 'bar', title = 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY CAT

Out[48]:  <Axes: title={'center': 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY CATEGORY'}, xlabel
          ='Category', ylabel='total amount spent'>
```

## TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY CATEGORY
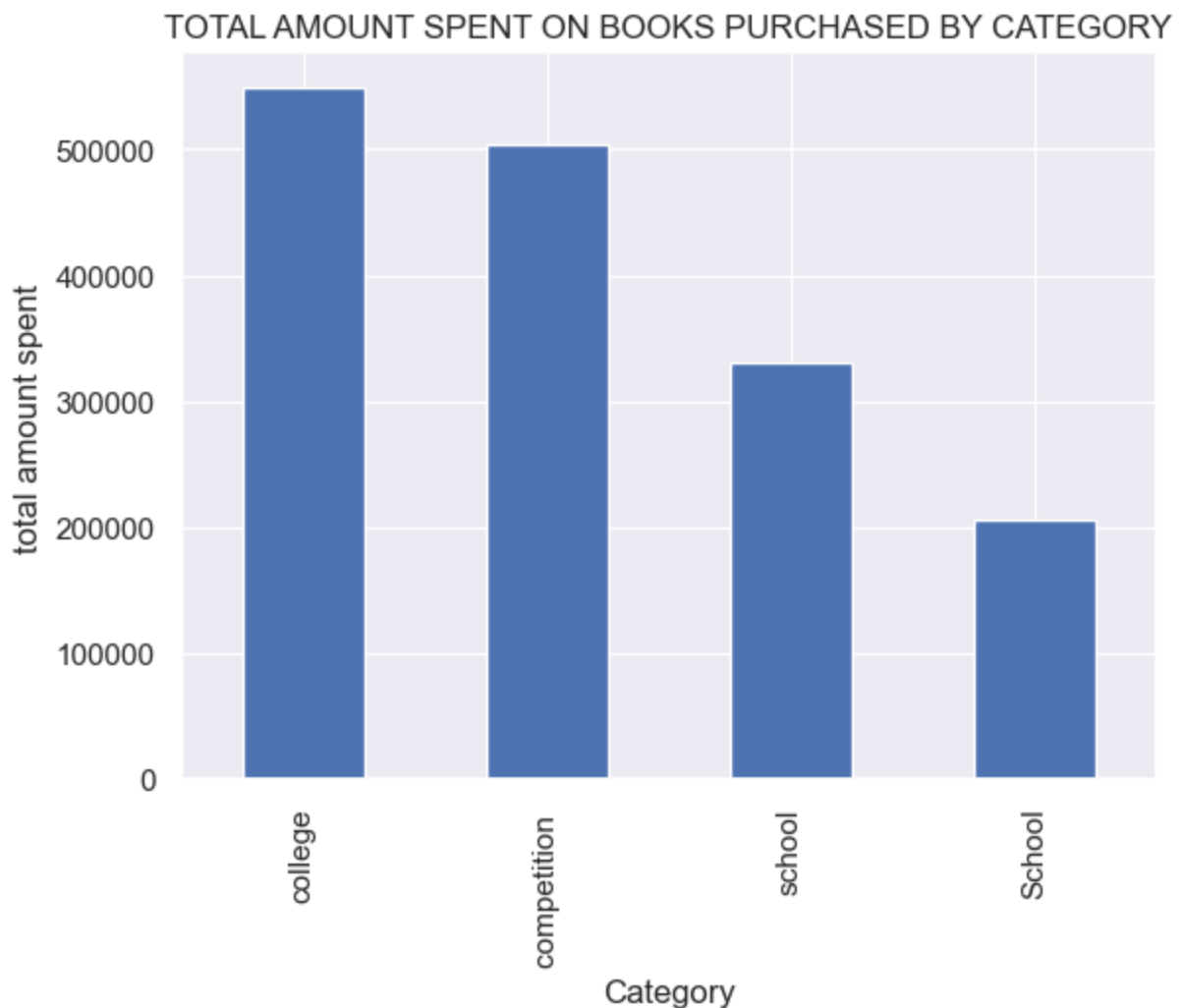

```
In [49]:  #total amount by Author
          top_author1= kaggle.groupby('Author')['Total_amount'].sum().sort_values(ascending = Fa
          top_author1
```

```
Out[49]:  Author
          NCERT                                     407290.39
          MAHAVEER JAIN,ANSHUL JAIN KANTI JAIN      113102.61
          preeti agarwal, r.s. sharma, vinod gupta   73107.00
          Preeti Agarwal                             50210.00
          N. Gregory Mankiw                          47661.00
          Geoffrey Leech                             46401.00
          DR.Â ALI AKHTER                            46215.00
          Â M.D. Raisinghania                        37681.00
          Dr.H.L.AHUJA                               32790.00
          Ahuja H.L.                                 29229.00
          Name: Total_amount, dtype: float64
```

```
In [50]:  top_author1.plot(kind = 'bar', title = 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY AUTHC
```

```
Out[50]:  <Axes: title={'center': 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY AUTHOR'}, xlabel='A
          uthor', ylabel='total amount spent'>
```

TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY AUTHOR
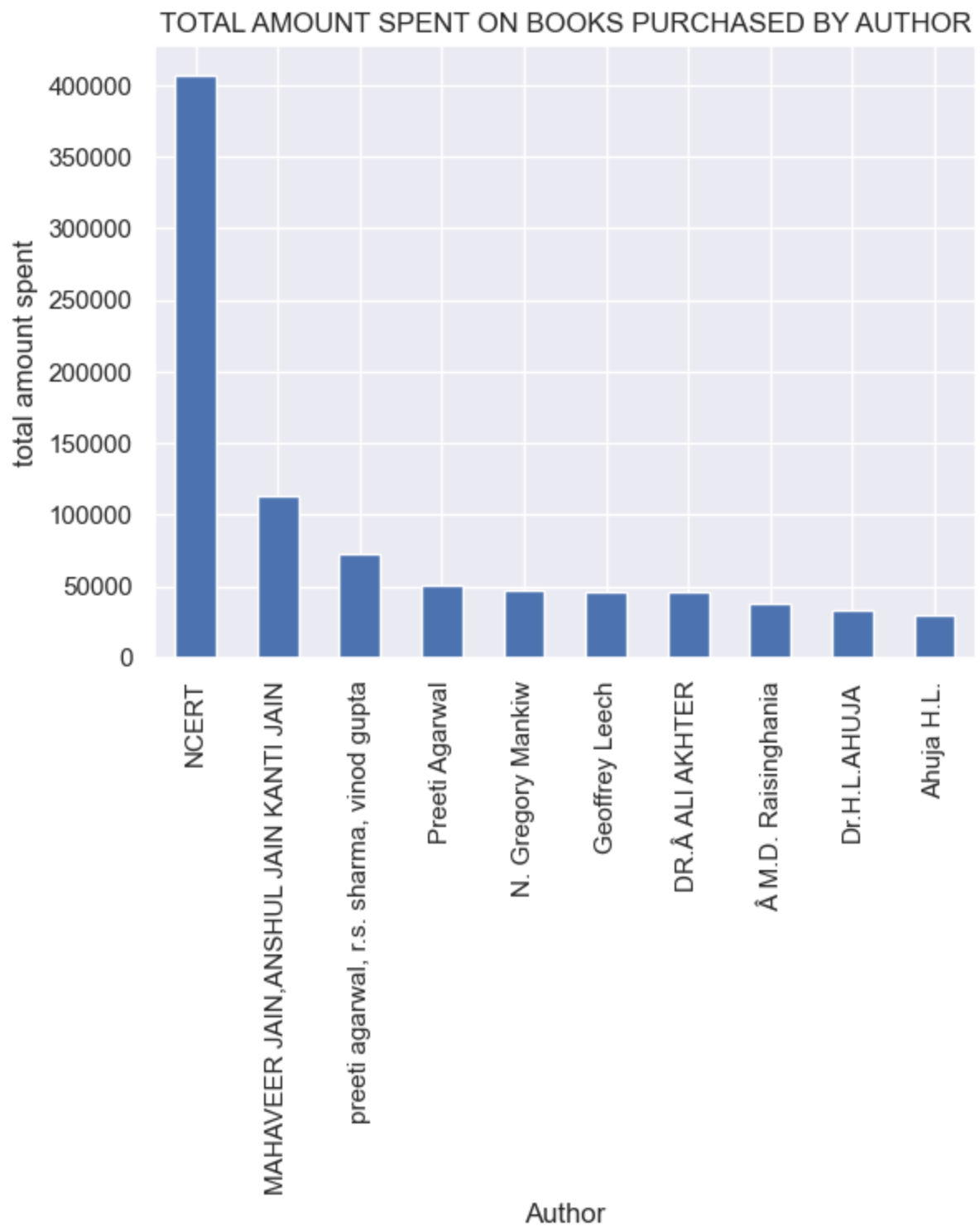


```
In [51]:  #total amount by publication
          top_publication1 = kaggle.groupby('Publication')['Total_amount'].sum().sort_values(asc
          top_publication1
```

Out[51]:
```
Publication
National Council of Education Research and Training          339217.64
VK Global Publications Private Limited                       116657.61
Vardhan                                                       73107.00
National Council of Education Research and Training (2014)    60728.37
S. Chand Publishing                                           55948.00
Manu Prakashan Private Limited                                52426.78
Arihant Publications                                          51510.00
Vardhan Publishers & Distributors                             50210.00
OpenStax                                                      47661.00
SASA PUBLICATION                                              46215.00
Name: Total_amount, dtype: float64
```

In [52]:
```python
top_publication1.plot(kind = 'bar', title = 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY
```

Out[52]:
```
<Axes: title={'center': 'TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY PUBLICATION'}, xlab
el='Publication', ylabel='total amount spent'>
```

TOTAL AMOUNT SPENT ON BOOKS PURCHASED BY PUBLICATION
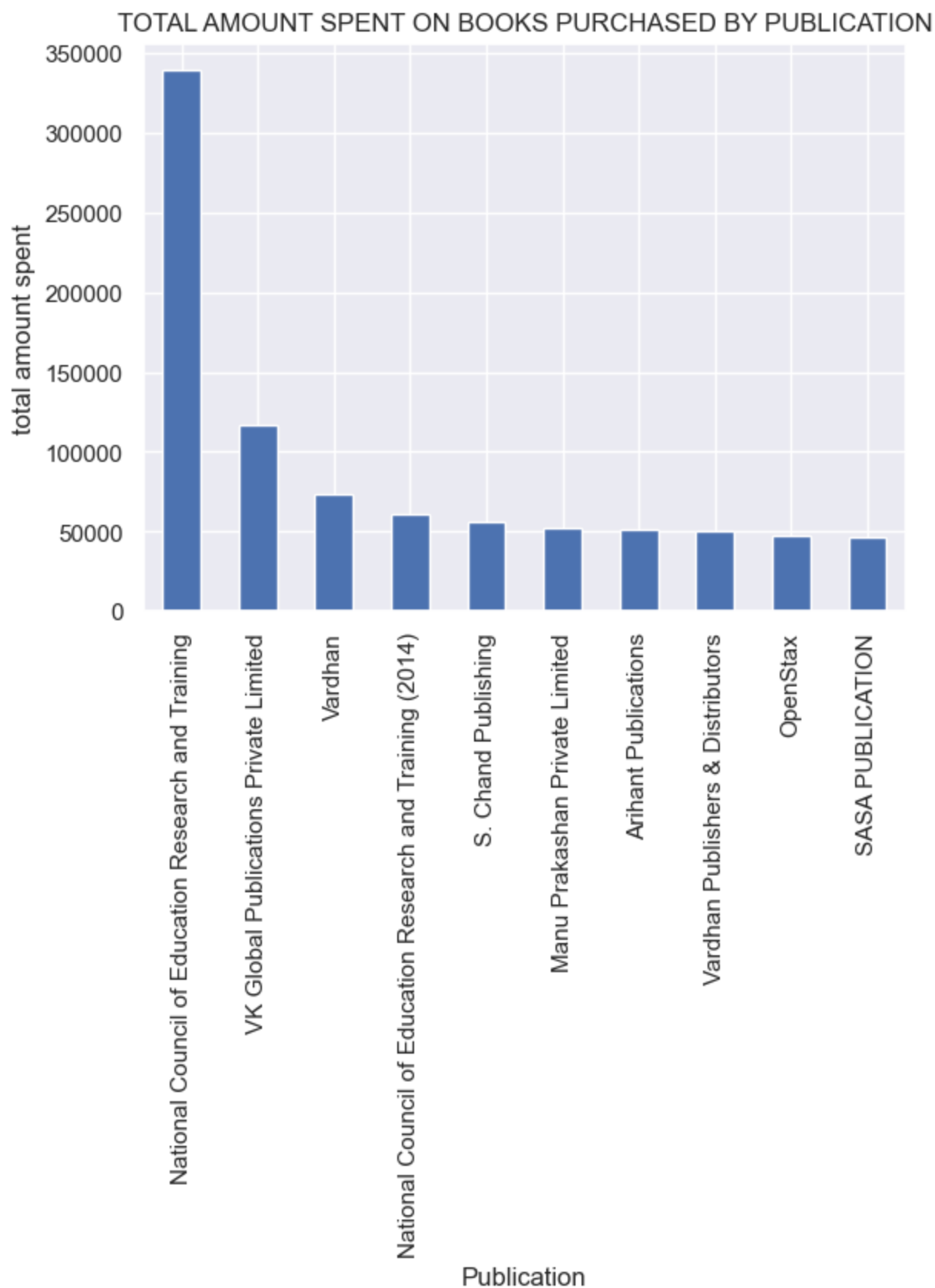

```
In [53]:  #total amount purchased by customer
          top_customer1 = kaggle.groupby('Customer_ID')['Total_amount'].sum().sort_values(ascenc
          top_customer1
```

Out[53]:
```
Customer_ID
B07FWWBYQ    7300.0
B07FWXZXT    5912.0
B07FWVUUU    5854.5
B07FWVUVY    5440.5
B07FWVVWQ    4578.0
B07FWVVRW    4332.0
B07FWVVYX    4057.5
B07FWVUSW    3914.0
B07FWVWWW    3815.0
B07FWXUTS    3808.0
Name: Total_amount, dtype: float64
```

In [54]:
```python
top_customer1.plot(kind = 'bar', title = 'TOP CUSTOMERS BY TOTAL AMOUNT SPENT', ylabel
```

Out[54]:
```
<Axes: title={'center': 'TOP CUSTOMERS BY TOTAL AMOUNT SPENT'}, xlabel='Customer_ID',
ylabel='total amount spent'>
```
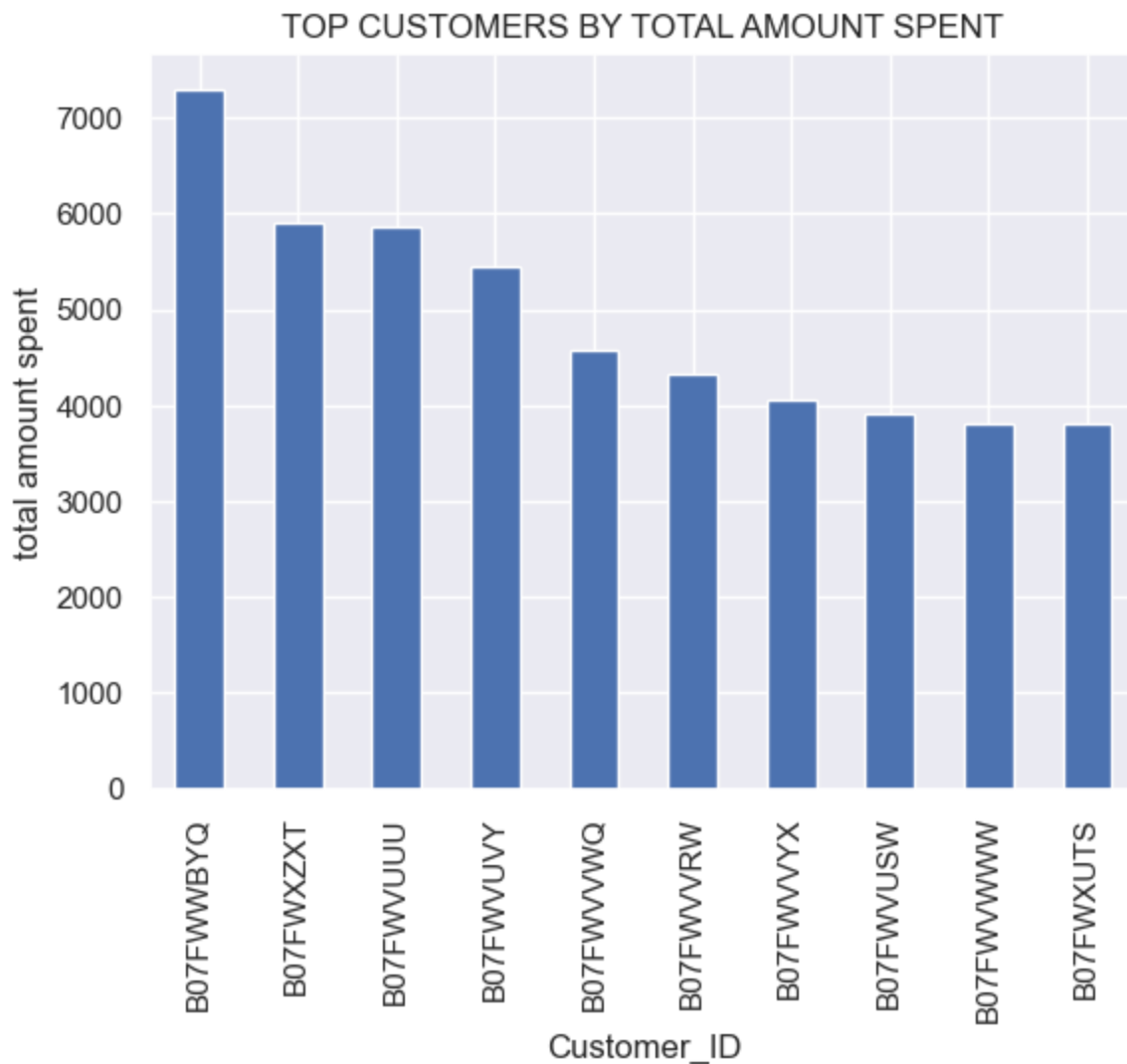

In [55]:
```python
#profit made by purchase date
top_date2 = kaggle.groupby('Purchase_Date')['Profit(INR)'].sum().sort_values(ascending
top_date2
```

```
Out[55]:   Purchase_Date
           2019-07-28    1796.230
           2019-06-01    1718.380
           2019-07-27    1643.224
           2018-04-09    1597.690
           2019-06-07    1569.521
           2018-05-13    1545.300
           2018-01-15    1400.850
           2018-03-29    1329.940
           2018-05-02    1276.880
           2018-02-25    1272.960
           Name: Profit(INR), dtype: float64
```

```python
In [56]:   top_date2.plot(kind = 'line', title = 'HIGHLY PROFITABLE PURCHASE_DATES', ylabel = 'Pr
```

```
Out[56]:   <Axes: title={'center': 'HIGHLY PROFITABLE PURCHASE_DATES'}, xlabel='Purchase_Date',
           ylabel='Profit(INR)'>
```
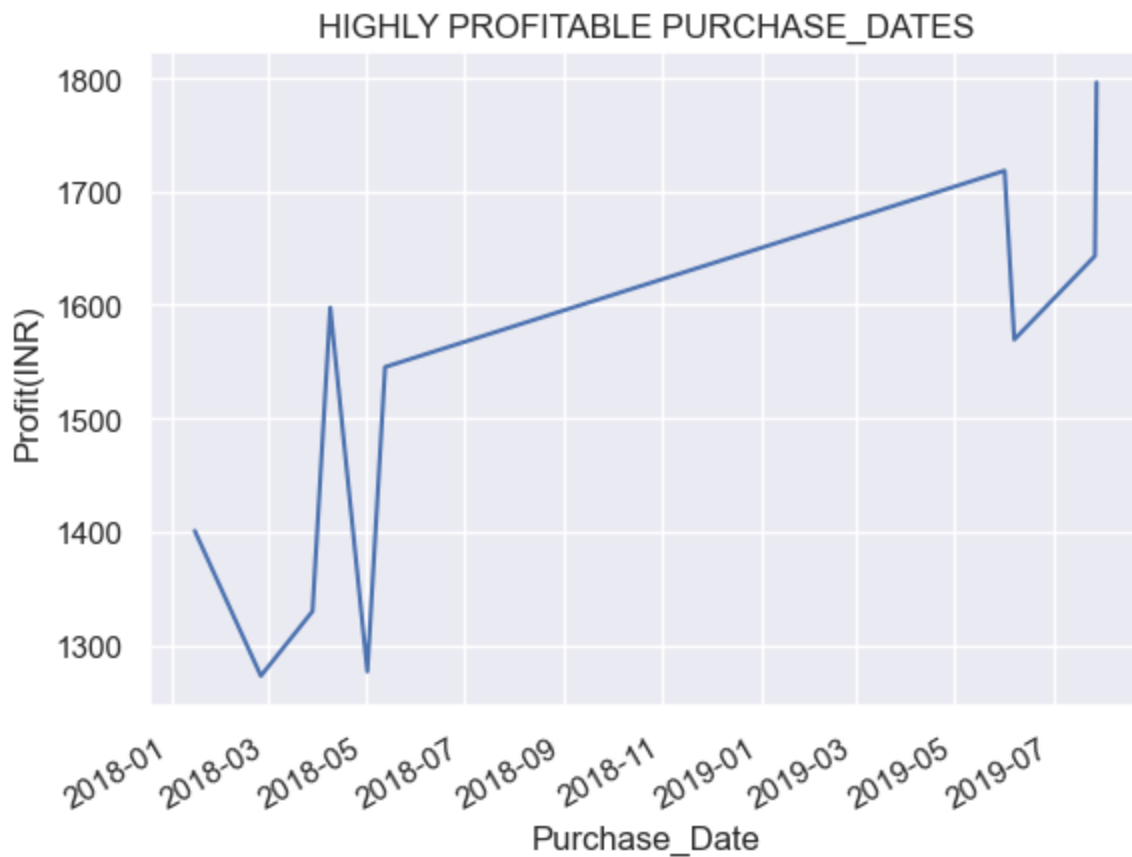


```python
In [57]:   #profit made by year
           top_year2 = kaggle.groupby('Year')['Profit(INR)'].sum().sort_values(ascending = False)
           top_year2
```

```
Out[57]:   Year
           2019    108358.4272
           2018     99780.7500
           Name: Profit(INR), dtype: float64
```

```python
In [58]:   top_year2.plot(kind = 'bar', title = 'PROFIT MADE BY YEARS', ylabel = 'Profit(INR)')
```

```
Out[58]:   <Axes: title={'center': 'PROFIT MADE BY YEARS'}, xlabel='Year', ylabel='Profit(INR)'>
```

PROFIT MADE BY YEARS


```
In [59]:   #profit made by quaters
           top_quater2= kaggle.groupby('Quater')['Profit(INR)'].sum().sort_values(ascending = Fal
           top_quater2
```

```
Out[59]:   Quater
           2    74210.4514
           3    55674.6278
           1    50339.2160
           4    27914.8820
           Name: Profit(INR), dtype: float64
```

```
In [60]:   top_quater2.plot(kind = 'bar', title = 'PROFIT MADE BY QUATER', ylabel = 'Profit(INR)'
```

```
Out[60]:   <Axes: title={'center': 'PROFIT MADE BY QUATER'}, xlabel='Quater', ylabel='Profit(IN
           R)'>
```
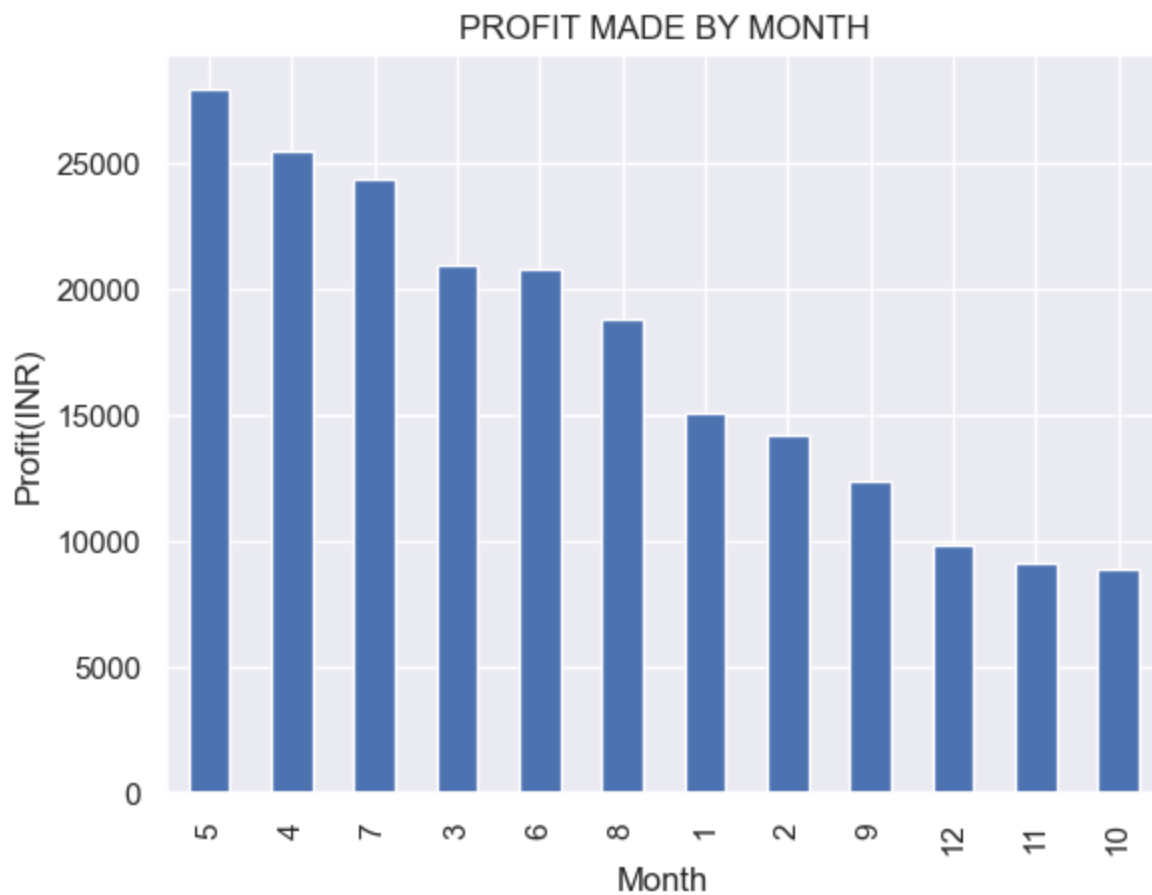
PROFIT MADE BY QUATER


```
In [61]:   #profit made on books by month
           top_month2= kaggle.groupby('Month')['Profit(INR)'].sum().sort_values(ascending = False
           top_month2
```

```
Out[61]:   Month
           5      27930.9130
           4      25460.2544
           7      24393.0564
           3      20976.3460
           6      20819.2840
           8      18861.5544
           1      15132.6200
           2      14230.2500
           9      12420.0170
           12      9857.4650
           11      9177.2400
           10      8880.1770
           Name: Profit(INR), dtype: float64
```

```
In [62]:   top_month2.plot(kind = 'bar', title = 'PROFIT MADE BY MONTH', ylabel = 'Profit(INR)')
```

```
Out[62]:   <Axes: title={'center': 'PROFIT MADE BY MONTH'}, xlabel='Month', ylabel='Profit(IN
           R)'>
```
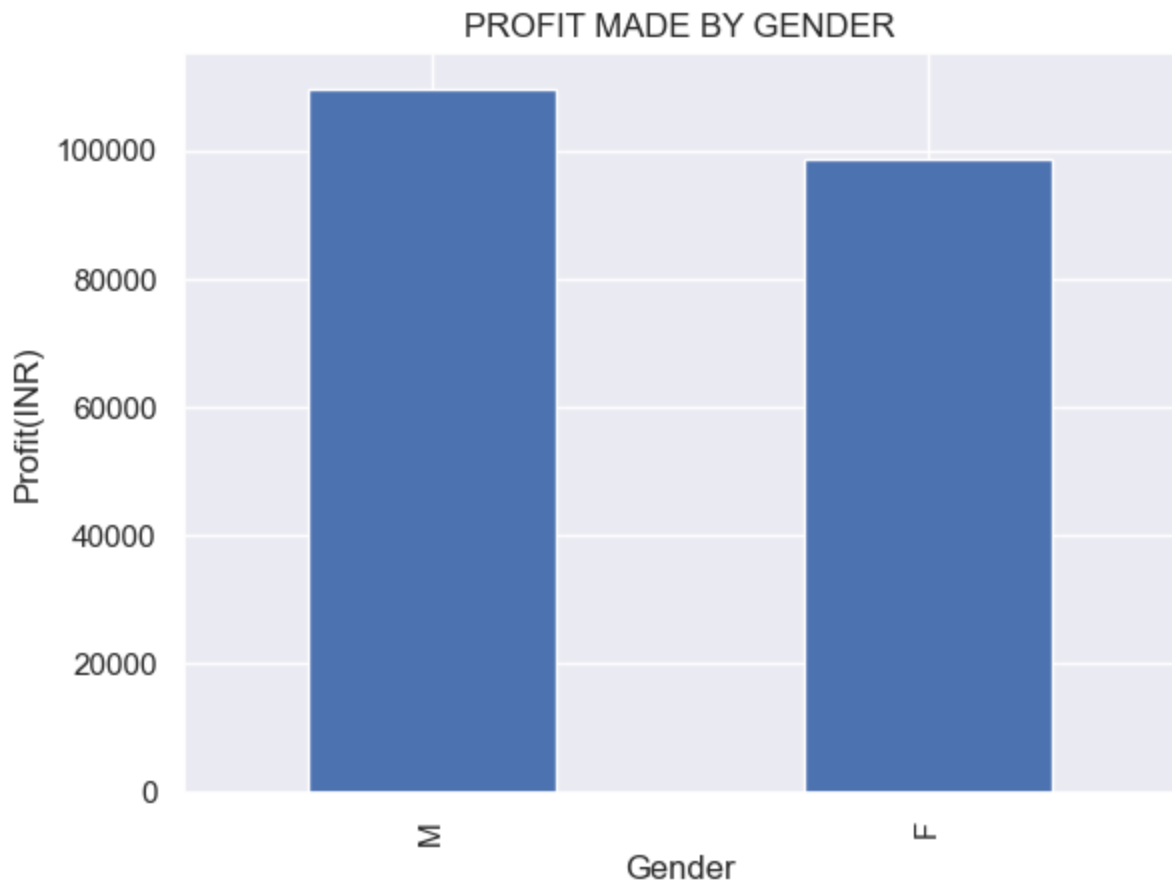
PROFIT MADE BY MONTH

In [63]:
```python
#profit made by gender
top_gender2= kaggle.groupby('Gender')['Profit(INR)'].sum().sort_values(ascending = Fal
top_gender2
```

Out[63]:
```
Gender
M     109645.3394
F      98493.8378
Name: Profit(INR), dtype: float64
```

In [64]:
```python
top_gender2.plot(kind = 'bar', title = 'PROFIT MADE BY GENDER', ylabel = 'Profit(INR)'
```

Out[64]:
```
<Axes: title={'center': 'PROFIT MADE BY GENDER'}, xlabel='Gender', ylabel='Profit(IN
R)'>
```
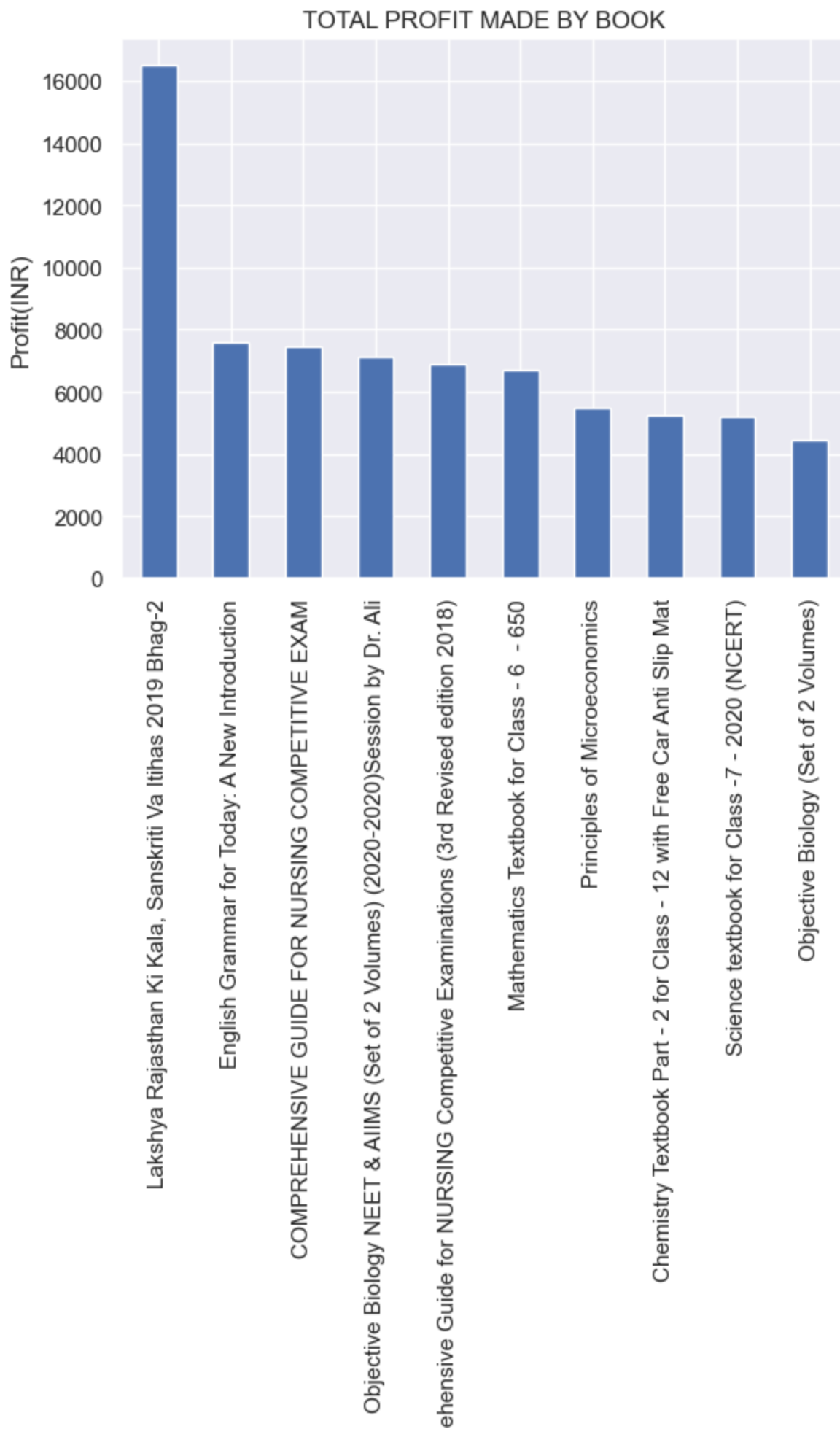
PROFIT MADE BY GENDER


```
In [65]:  #total profit made by product(book) name
          top_product2= kaggle.groupby('Product_Name')['Profit(INR)'].sum().sort_values(ascendir
          top_product2
```

```
Out[65]:  Product_Name
          Lakshya Rajasthan Ki Kala, Sanskriti Va Itihas 2019 Bhag-2
          16518.90
          English Grammar for Today: A New Introduction
          7595.28
          COMPREHENSIVE GUIDE FOR NURSING COMPETITIVE EXAM
          7441.92
          Objective Biology NEET & AIIMS (Set of 2 Volumes) (2020-2020)Session by Dr. Ali
          7125.72
          Vardhan Comprehensive Guide for NURSING Competitive Examinations (3rd Revised edition
          2018)      6879.60
          Mathematics Textbook for Class - 6  - 650
          6720.30
          Principles of Microeconomics
          5478.33
          Chemistry Textbook Part - 2 for Class - 12 with Free Car Anti Slip Mat
          5253.12
          Science textbook for Class -7 - 2020 (NCERT)
          5229.00
          Objective Biology (Set of 2 Volumes)
          4467.60
          Name: Profit(INR), dtype: float64
```

```
In [66]:  top_product2.plot(kind = 'bar', title = 'TOTAL PROFIT MADE BY BOOK', ylabel = 'Profit(
```

```
Out[66]:  <Axes: title={'center': 'TOTAL PROFIT MADE BY BOOK'}, xlabel='Product_Name', ylabel
          ='Profit(INR)'>
```

TOTAL PROFIT MADE BY BOOK

Bar chart titled "TOTAL PROFIT MADE BY BOOK" with y-axis labeled "Profit(INR)" ranging from 0 to 16000. Books along x-axis: Lakshya Rajasthan Ki Kala, Sanskriti Va Itihas 2019 Bhag-2 (~16500); English Grammar for Today: A New Introduction (~7600); COMPREHENSIVE GUIDE FOR NURSING COMPETITIVE EXAM (~7400); Objective Biology NEET & AIIMS (Set of 2 Volumes) (2020-2020)Session by Dr. Ali (~7100); ehensive Guide for NURSING Competitive Examinations (3rd Revised edition 2018) (~6900); Mathematics Textbook for Class - 6 - 650 (~6700); Principles of Microeconomics (~5500); Chemistry Textbook Part - 2 for Class - 12 with Free Car Anti Slip Mat (~5300); Science textbook for Class -7 - 2020 (NCERT) (~5200); Objective Biology (Set of 2 Volumes) (~4500).

Vardhan Compr

Product_Name

In [67]:
```python
#Total profit made by ship city
top_city2= kaggle.groupby('Ship_City')['Profit(INR)'].sum().sort_values(ascending = Fa
top_city2
```

Out[67]:
```
Ship_City
NEW DELHI     13412.3108
BENGALURU      9154.0462
MUMBAI         8010.1048
PUNE           6379.7522
JAIPUR         6139.6410
AHMEDABAD      4591.1978
CHENNAI        3384.9018
GURUGRAM       3294.1374
HYDERABAD      3204.1424
LUCKNOW        2560.1032
Name: Profit(INR), dtype: float64
```

In [68]:
```python
top_city2.plot(kind = 'bar', title = 'TOTAL PROFIT MADE BY EACH CITY', ylabel = 'Profi
```

Out[68]:
```
<Axes: title={'center': 'TOTAL PROFIT MADE BY EACH CITY'}, xlabel='Ship_City', ylabel
='Profit(INR)'>
```

TOTAL PROFIT MADE BY EACH CITY
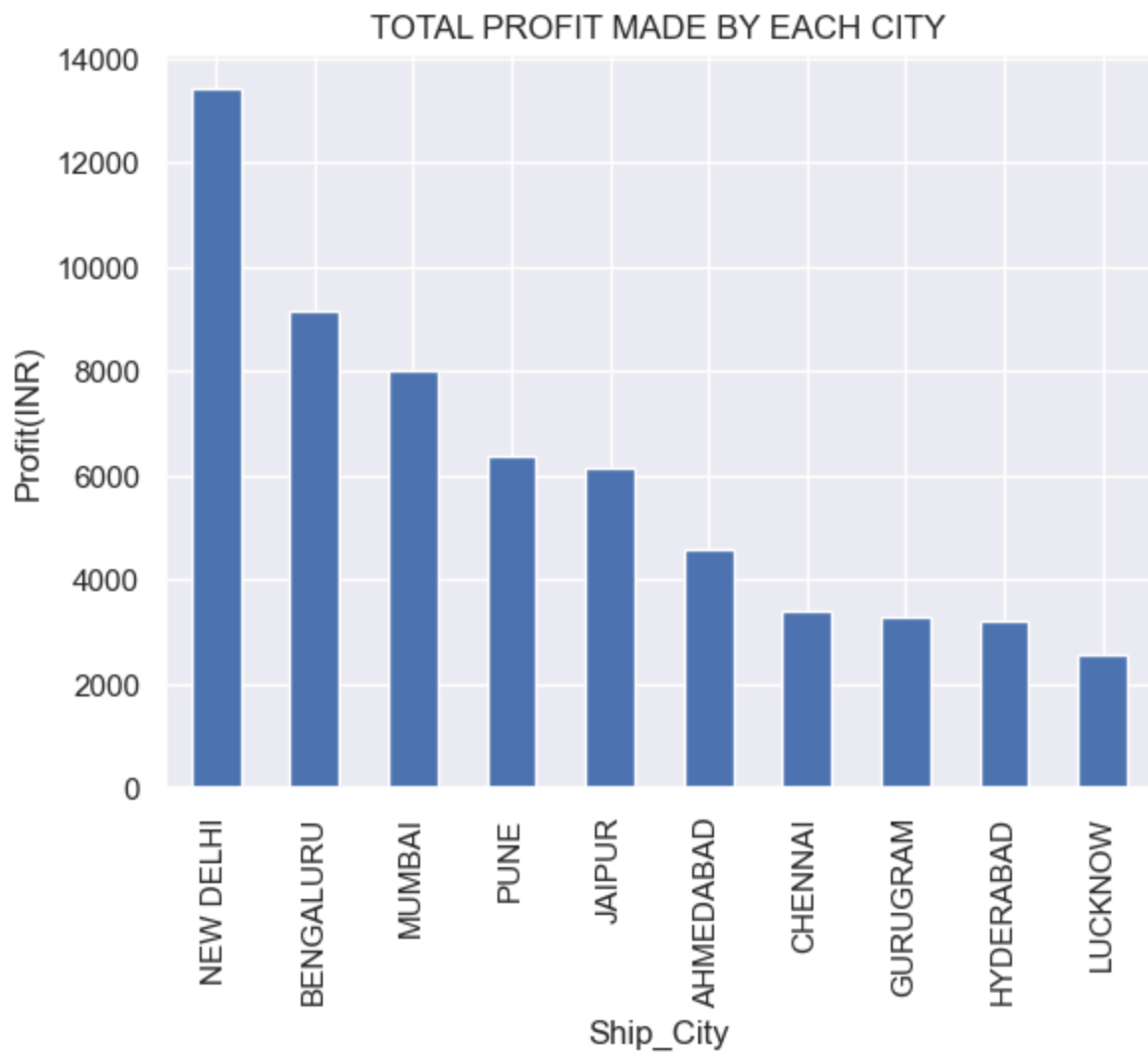

```
In [69]:  #quantity by ship state
          top_state2= kaggle.groupby('Ship_State')['Profit(INR)'].sum().sort_values(ascending =
          top_state2
```

```
Out[69]:  Ship_State
          MAHARASHTRA      37107.7466
          RAJASTHAN        21266.4178
          UTTAR PRADESH    18501.0892
          DELHI            15605.3368
          GUJARAT          14835.4036
          KARNATAKA        13992.6486
          TAMIL NADU       11599.7396
          HARYANA          10641.8128
          WEST BENGAL       8054.9518
          TELANGANA         5208.7562
          Name: Profit(INR), dtype: float64
```

```
In [70]:  top_state2.plot(kind = 'bar', title = 'TOTAL PROFIT MADE BY EACH STATE', ylabel = 'Prc
```

```
Out[70]:  <Axes: title={'center': 'TOTAL PROFIT MADE BY EACH STATE'}, xlabel='Ship_State', ylab
          el='Profit(INR)'>
```

TOTAL PROFIT MADE BY EACH STATE
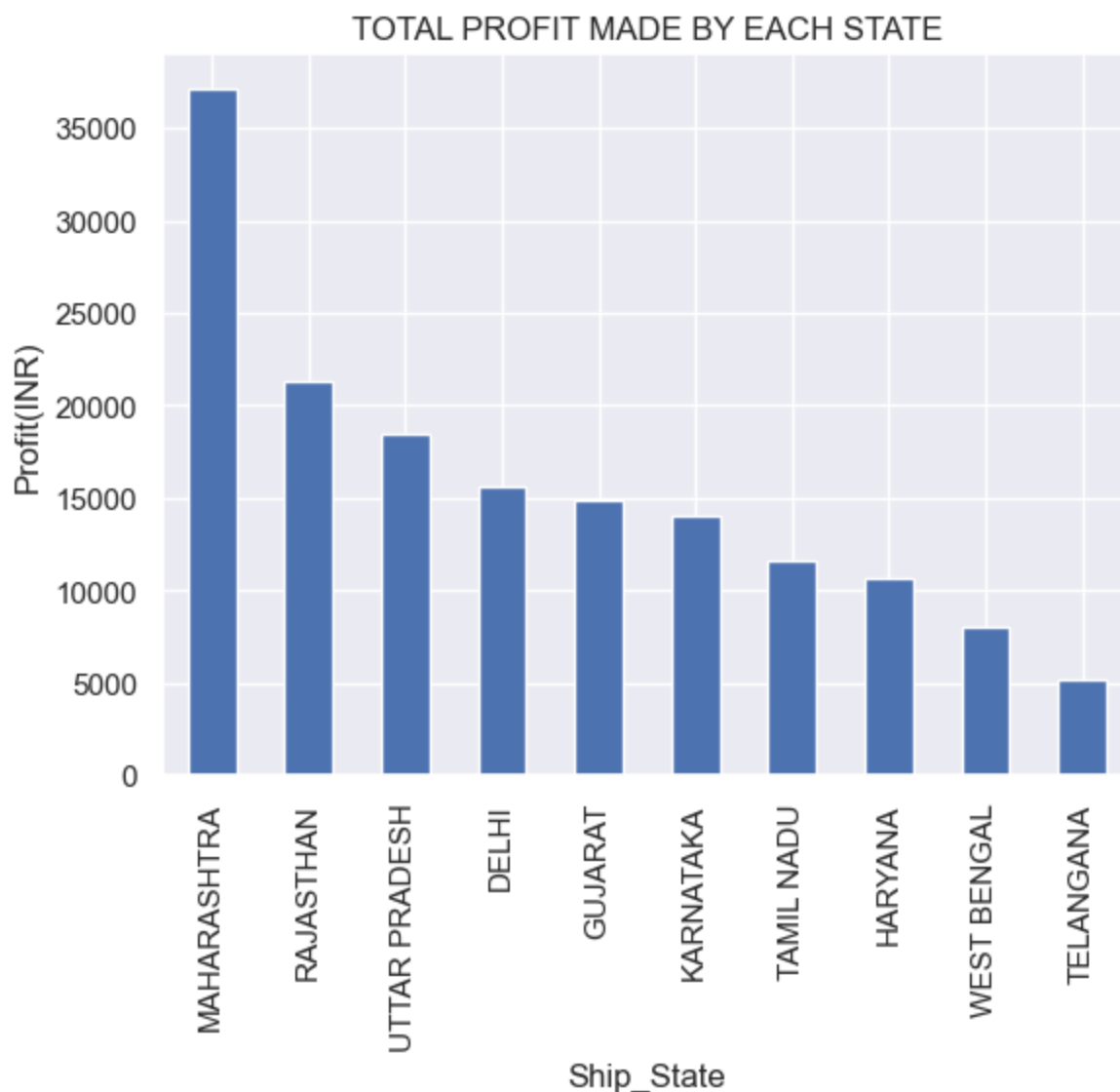
In [71]:
```python
#total profit by category
top_category2= kaggle.groupby('Category')['Profit(INR)'].sum().sort_values(ascending =
top_category2
```

Out[71]:
```
Category
competition     74385.5400
college         62676.4450
school          41701.5522
School          29375.6400
Name: Profit(INR), dtype: float64
```

In [72]:
```python
top_category2.plot(kind = 'bar', title = 'TOTAL PROFIT MADE BY CATEGORY', ylabel = 'Pr
```

Out[72]:
```
<Axes: title={'center': 'TOTAL PROFIT MADE BY CATEGORY'}, xlabel='Category', ylabel
='Profit(INR)'>
```

TOTAL PROFIT MADE BY CATEGORY
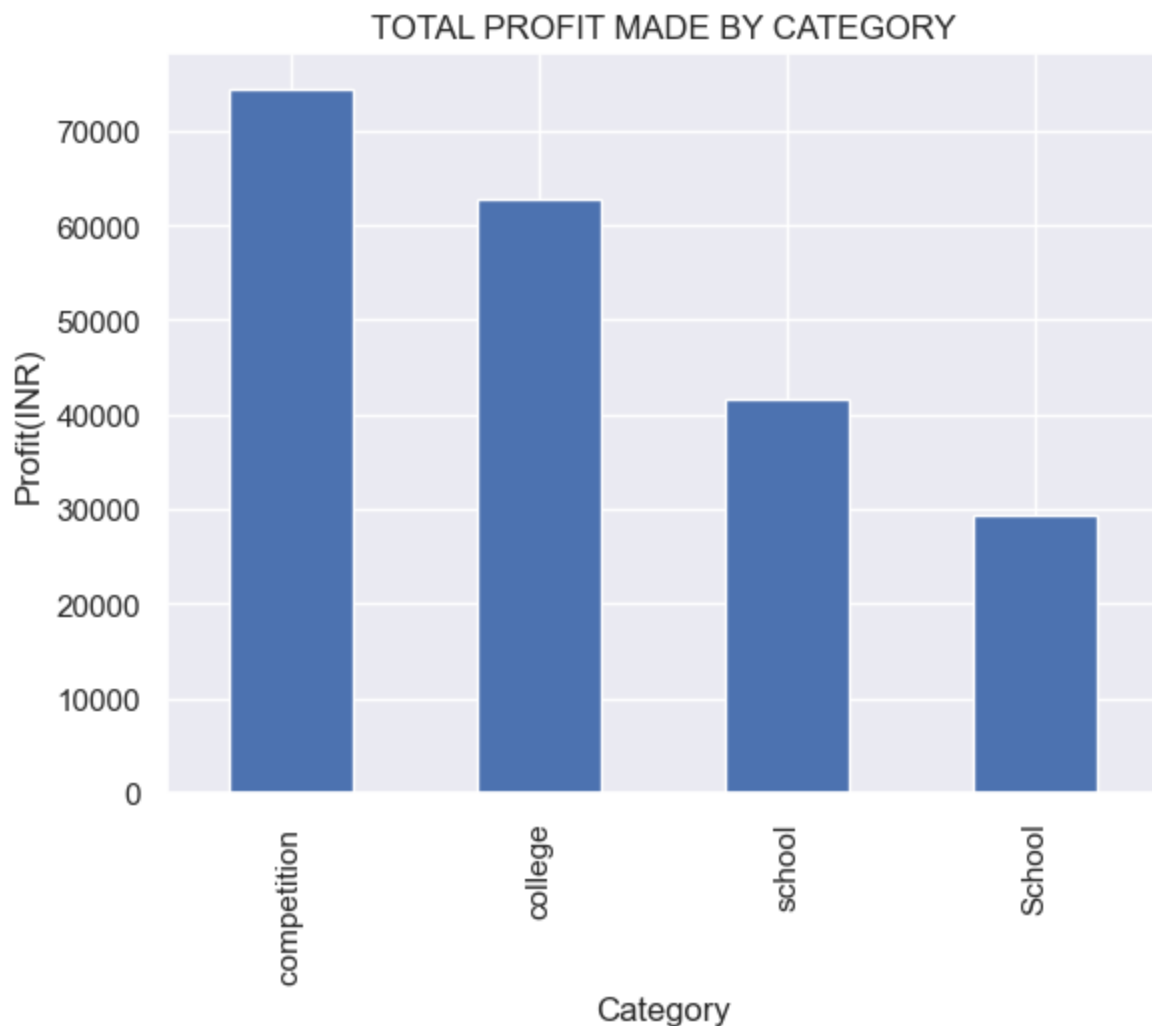


```
In [73]:  #total profit by Author
          top_author2= kaggle.groupby('Author')['Profit(INR)'].sum().sort_values(ascending = Fal
          top_author2
```

```
Out[73]:  Author
          NCERT                                        51693.7032
          MAHAVEER JAIN,ANSHUL JAIN KANTI JAIN         16518.9000
          preeti agarwal, r.s. sharma, vinod gupta      8626.8000
          Geoffrey Leech                                7675.9200
          Preeti Agarwal                                7441.9200
          DR.Â ALI AKHTER                               7125.7200
          N. Gregory Mankiw                             5478.3300
          A. Biswas, S. Biswas                          4467.6000
          Â M.D. Raisinghania                           4324.1900
          Kanti Jain                                    3916.8000
          Name: Profit(INR), dtype: float64
```

```
In [74]:  top_author2.plot(kind = 'bar', title = 'TOTAL PROFIT MADE BY AUTHOR', ylabel = 'Profit
```

```
Out[74]:  <Axes: title={'center': 'TOTAL PROFIT MADE BY AUTHOR'}, xlabel='Author', ylabel='Prof
          it(INR)'>
```

TOTAL PROFIT MADE BY AUTHOR
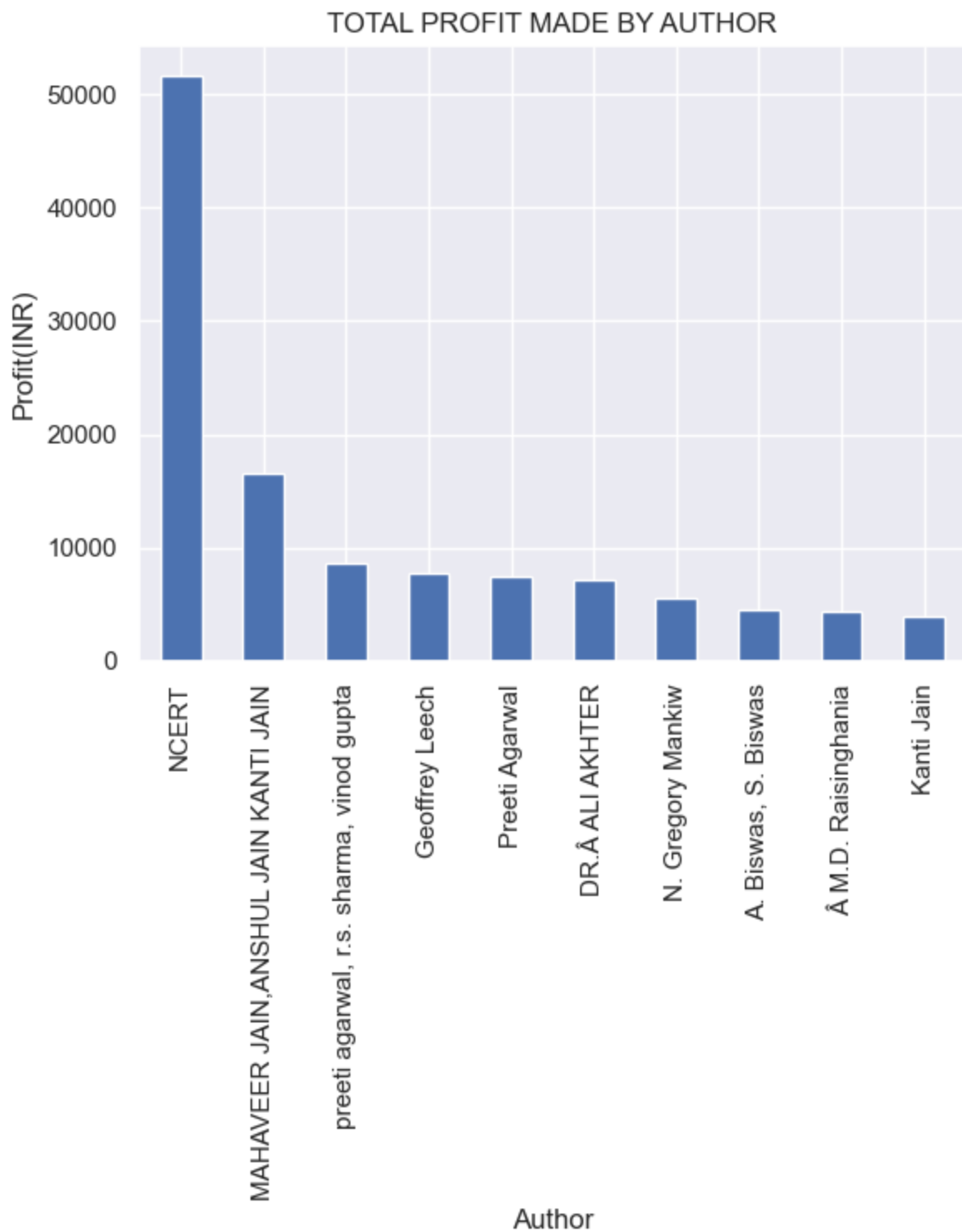
In [75]:
```python
#total profit by publication
top_publication2 = kaggle.groupby('Publication')['Profit(INR)'].sum().sort_values(asce
top_publication2
```

Out[75]:
```
Publication
National Council of Education Research and Training            42842.7432
VK Global Publications Private Limited                        17021.1000
Vardhan                                                        8626.8000
National Council of Education Research and Training (2014)     8028.9000
Palgrave Macmillan                                             7595.2800
Vardhan Publishers & Distributors                             7441.9200
Manu Prakashan Private Limited                                 7425.6000
Arihant Publications                                           7328.3900
SASA PUBLICATION                                               7125.7200
S. Chand Publishing                                            6428.7300
Name: Profit(INR), dtype: float64
```

In [76]: `top_publication2.plot(kind = 'bar', title = 'TOTAL PROFIT MADE BY PUBLICATION', ylabel`

Out[76]: `<Axes: title={'center': 'TOTAL PROFIT MADE BY PUBLICATION'}, xlabel='Publication', ylabel='Profit(INR)'>`

TOTAL PROFIT MADE BY PUBLICATION
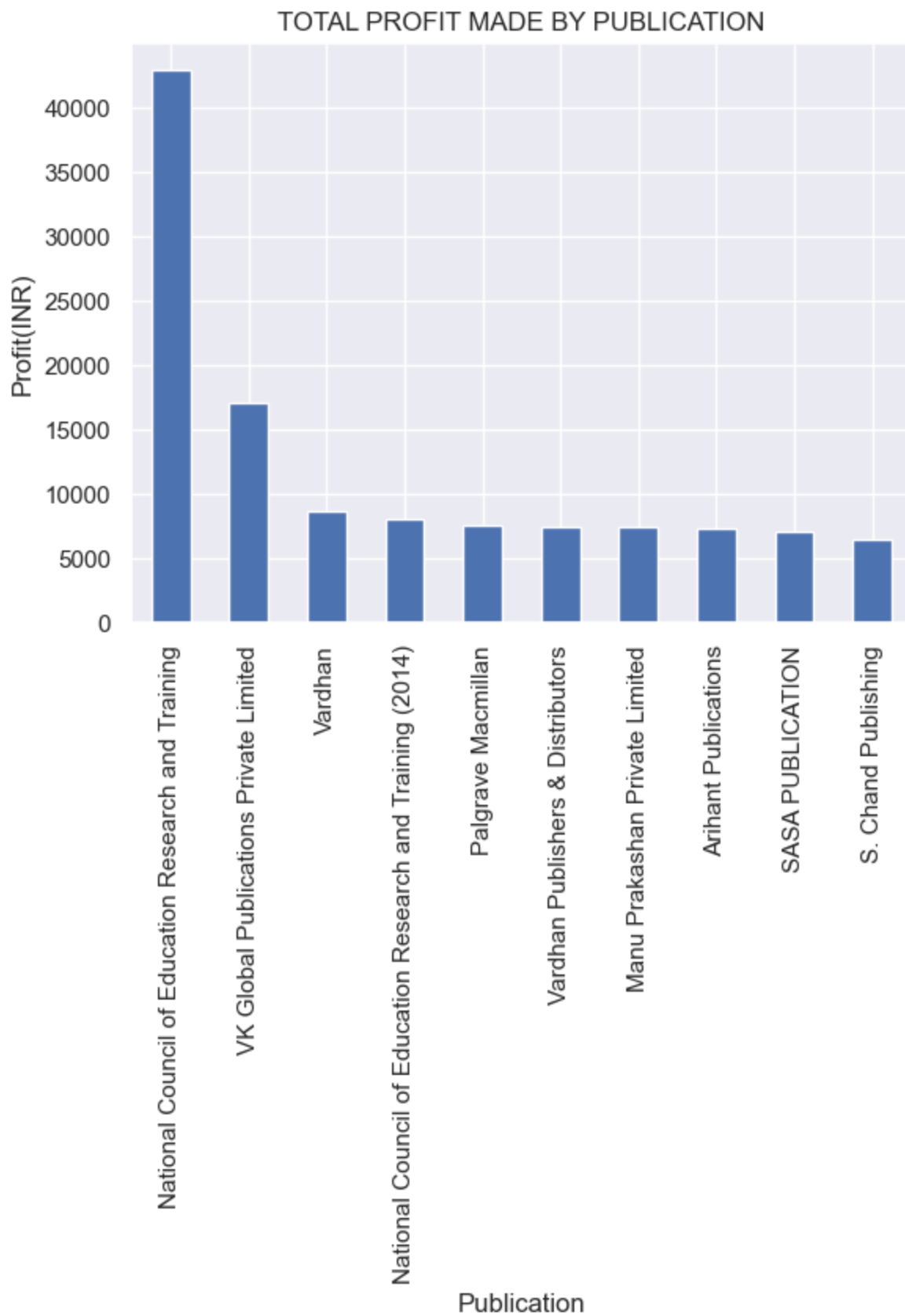


```
In [77]:  #total profit by customer
          top_customer2 = kaggle.groupby('Customer_ID')['Profit(INR)'].sum().sort_values(ascendi
          top_customer2
```

Out[77]:
```
Customer_ID
B07FWXZXT    1035.720
B07FWWBYQ     838.500
B07FWVUUU     722.105
B07FWVUVY     708.765
B07FWVVWQ     698.010
B07FWVVUV     551.391
B07FWVVRW     534.720
B07FWVUSV     532.080
B07FWVUSW     513.390
B07FWWXXS     508.980
Name: Profit(INR), dtype: float64
```

In [78]:
```python
top_customer2.plot(kind = 'bar', title = 'TOTAL PROFIT MADE BY CUSTOMER', ylabel = 'Pr
```

Out[78]:
```
<Axes: title={'center': 'TOTAL PROFIT MADE BY CUSTOMER'}, xlabel='Customer_ID', ylabe
l='Profit(INR)'>
```
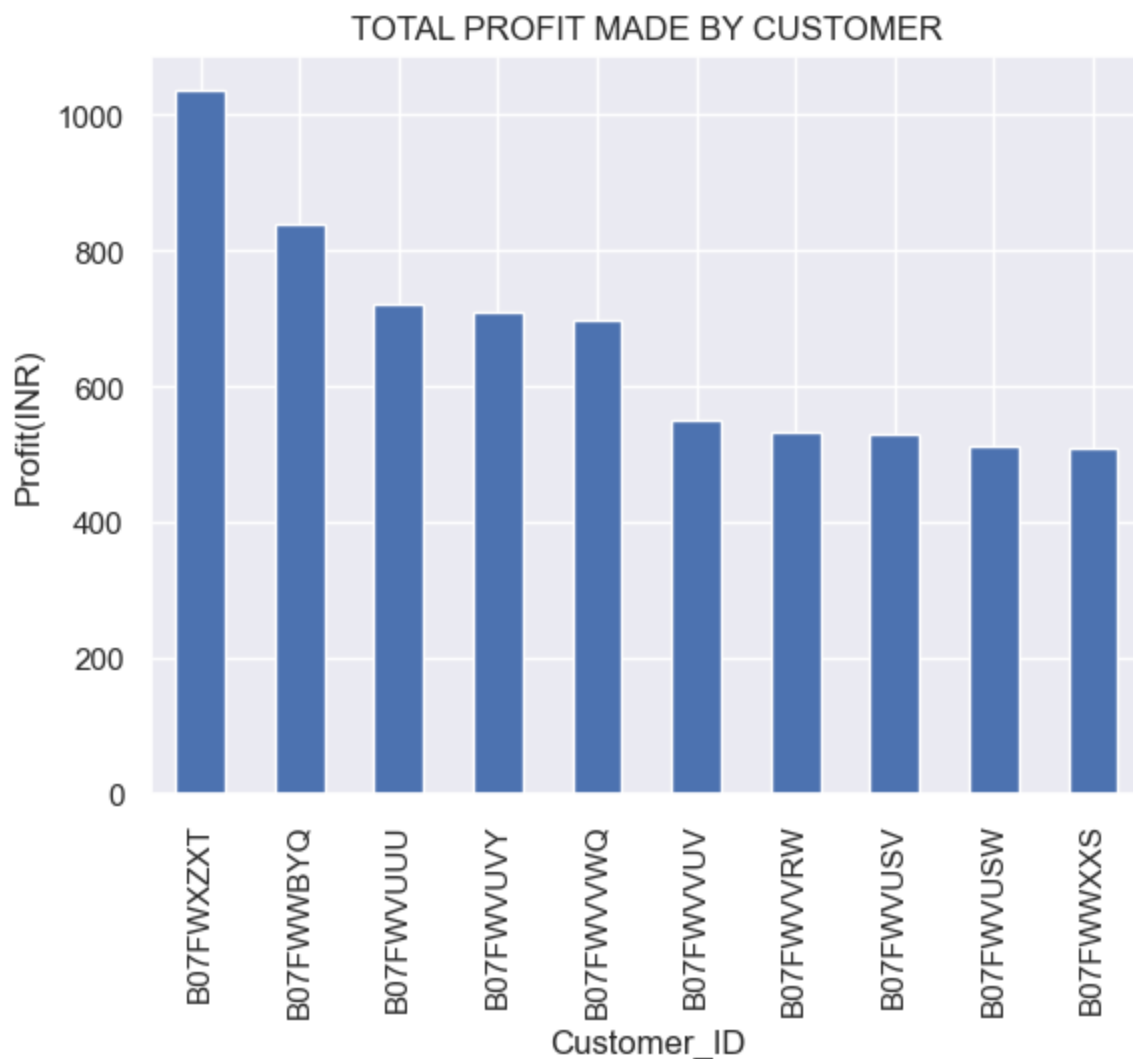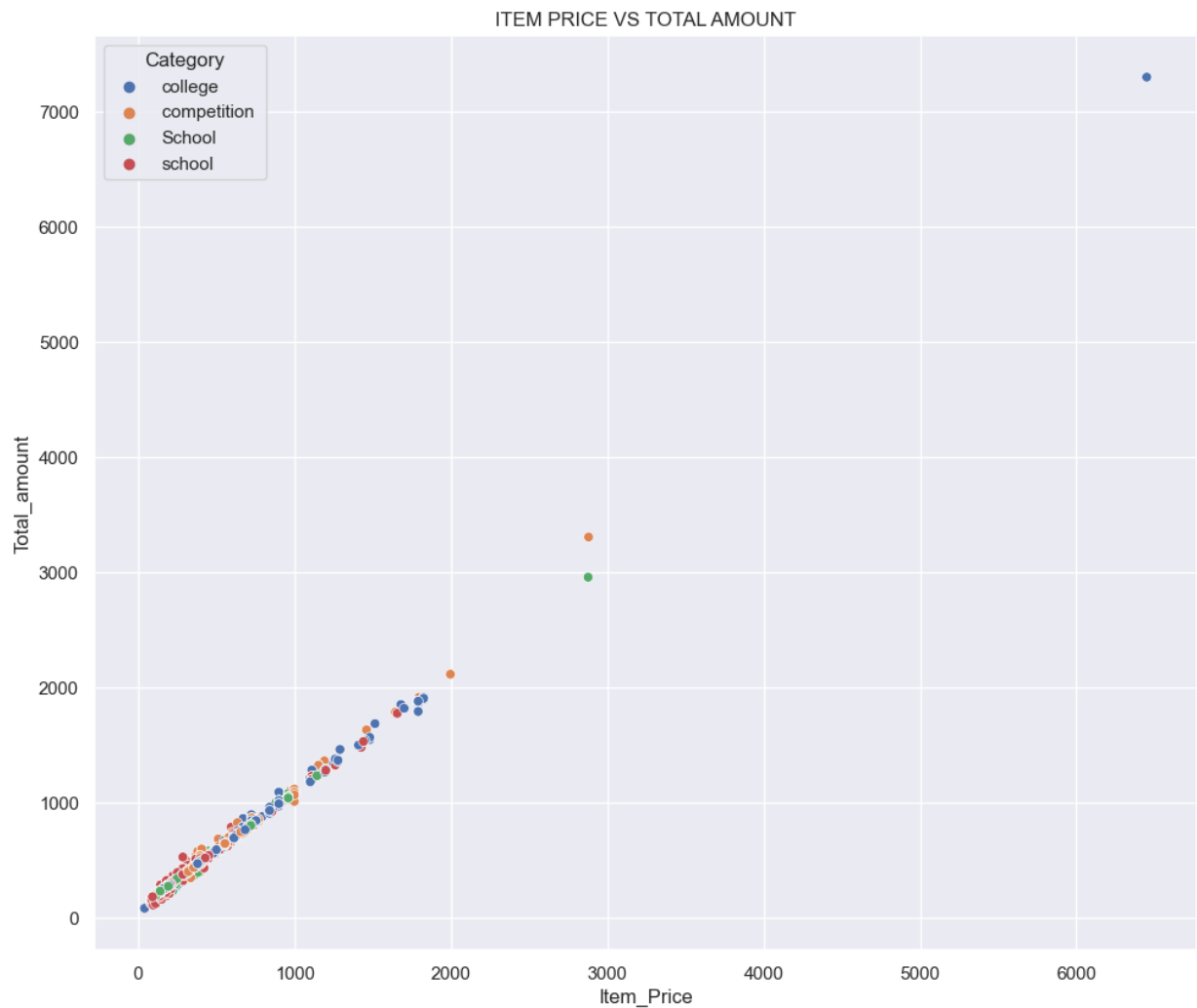

In [79]:
```python
plt.figure(figsize=(12,10))
axis = sns.scatterplot(x="Item_Price", y="Shipping_Price", data= kaggle, hue ='Categor
plt.title('ITEM PRICE VS SHIPPING PRICE')
```

Out[79]:
```
Text(0.5, 1.0, 'ITEM PRICE VS SHIPPING PRICE')
```

ITEM PRICE VS SHIPPING PRICE
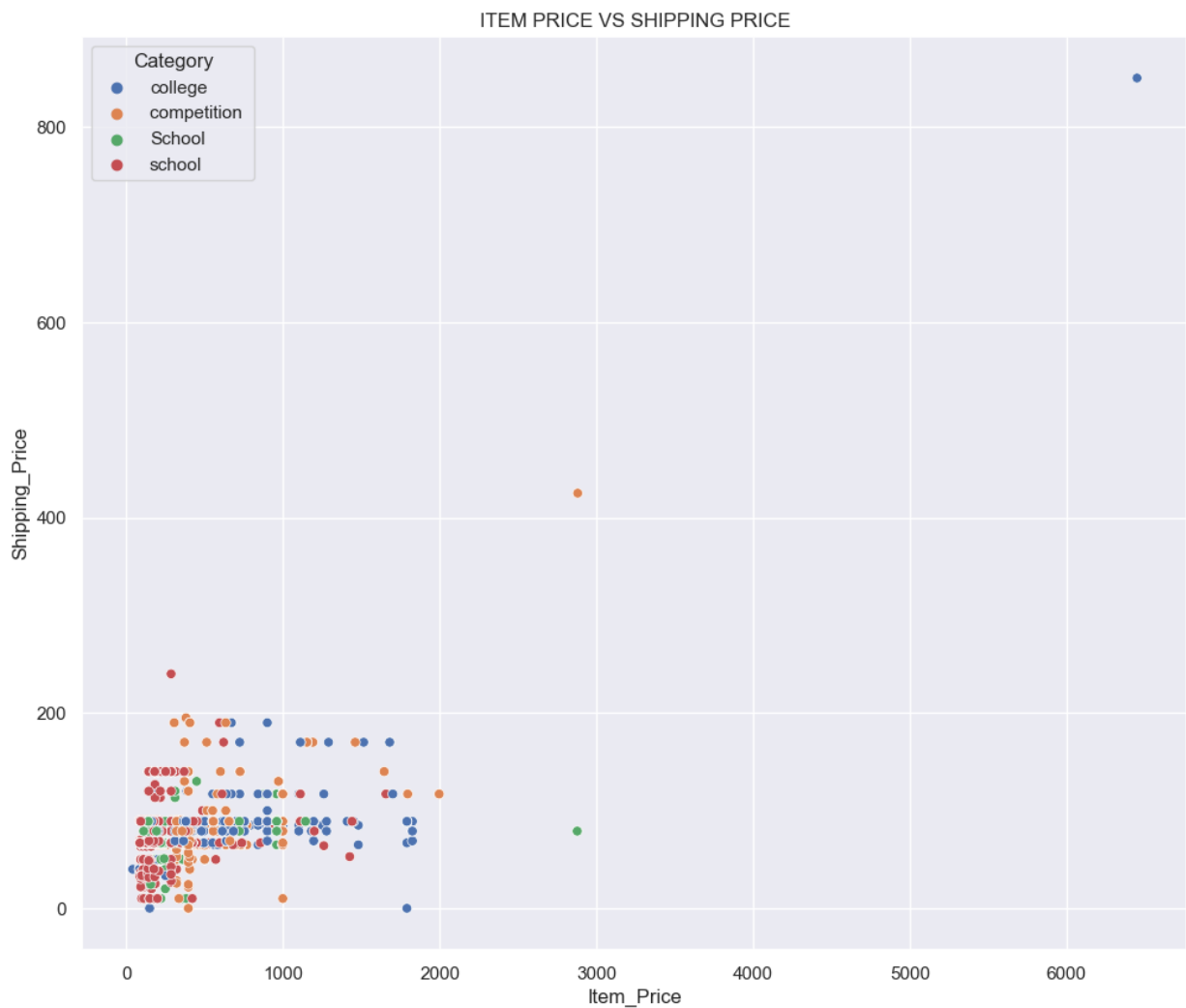

```
In [80]:  plt.figure(figsize=(12,10))
          axis = sns.scatterplot(x="Item_Price", y="Total_amount", data= kaggle, hue ='Category'
          plt.title('ITEM PRICE VS TOTAL AMOUNT')
```

Out[80]: Text(0.5, 1.0, 'ITEM PRICE VS TOTAL AMOUNT')

ITEM PRICE VS TOTAL AMOUNT


```
In [81]:  plt.figure(figsize=(12,10))
          axis = sns.scatterplot(x="Item_Price", y="Profit(INR)", data= kaggle, hue ='Category'
          plt.title('ITEM PRICE VS PROFIT(INR)')
```

Out[81]: Text(0.5, 1.0, 'ITEM PRICE VS PROFIT(INR)')
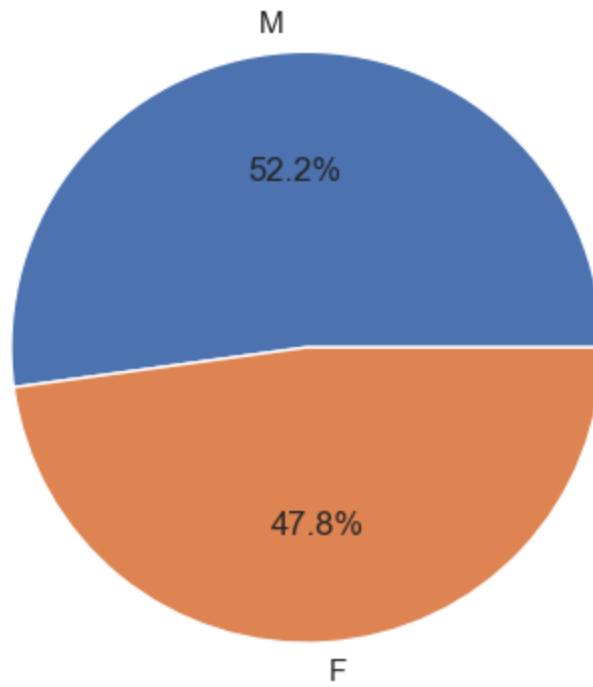
ITEM PRICE VS PROFIT(INR)
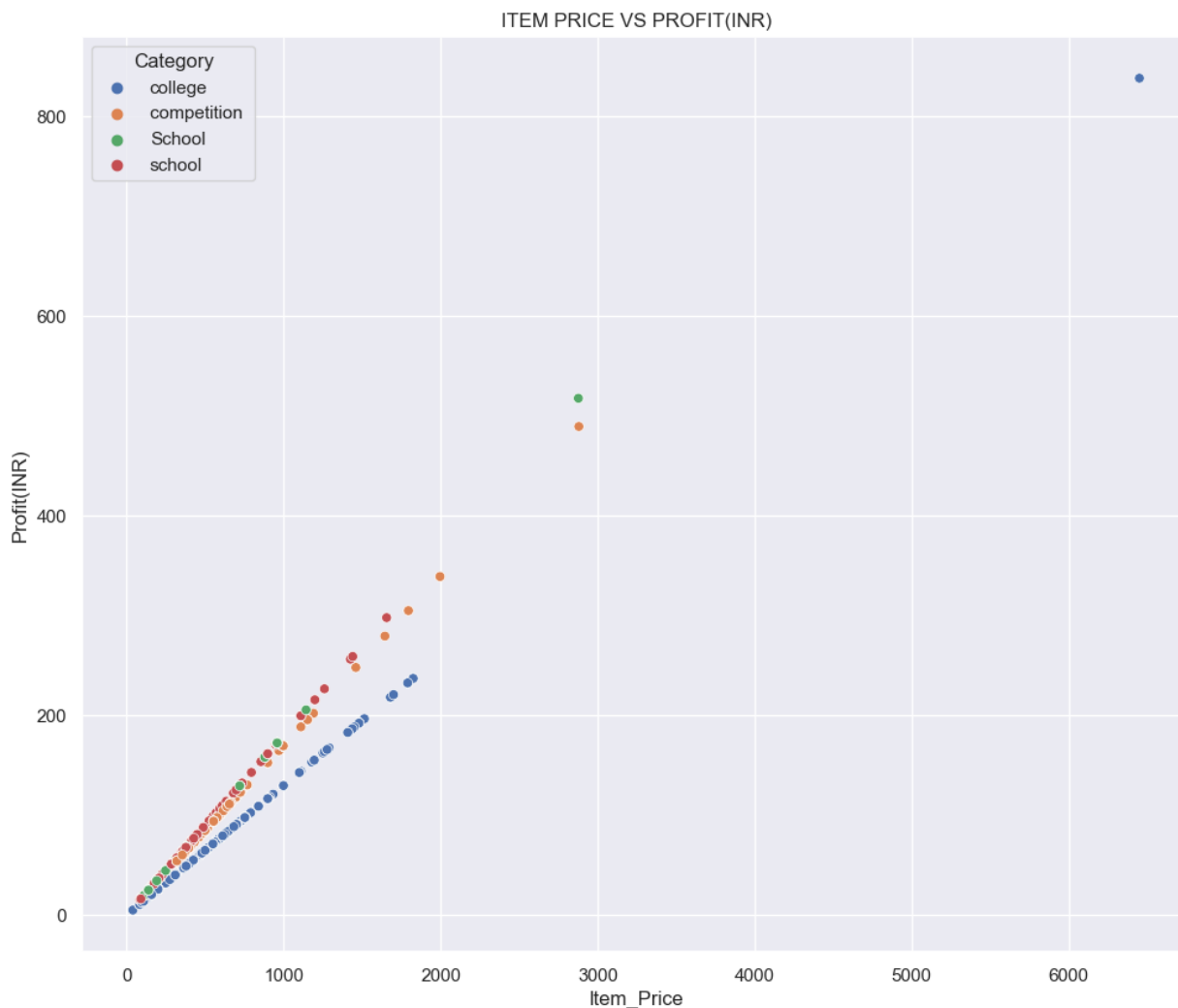

```
In [82]:   #Total_amount spent by gender
           top_gender3 = kaggle.groupby('Gender')['Gender'].count().sort_values(ascending = False
           top_gender3
```

```
Out[82]:   Gender
           M    1815
           F    1665
           Name: Gender, dtype: int64
```

```
In [83]:   top_gender3.plot.pie(autopct = '%1.1f%%')
           plt.title("GENDER DISTRIBUTION")
           plt.ylabel('')
```

```
Out[83]:   Text(0, 0.5, '')
```

GENDER DISTRIBUTION

M

52.2%

47.8%

F

In [ ]: