Data Analytic with TATA
by RStudio

Pakakorn Kaeoluan

2023/7/4

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.2      v readr     2.1.4
## v forcats   1.0.0      v stringr   1.5.0
## v ggplot2   3.4.2      v tibble    3.2.1
## v lubridate 1.9.2      v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
##
##     'rvest'
##
##
## The following object is masked from 'package:readr':
##
##     guess_encoding
```

1.Import Data

```
## # A tibble: 541,909 x 8
##    InvoiceNo StockCode Description      Quantity InvoiceDate         UnitPrice
##    <chr>     <chr>     <chr>               <dbl> <dttm>                  <dbl>
##  1 536365    85123A    WHITE HANGING HEA~      6 2010-12-01 08:26:00      2.55
##  2 536365    71053     WHITE METAL LANTE~      6 2010-12-01 08:26:00      3.39
##  3 536365    84406B    CREAM CUPID HEART~      8 2010-12-01 08:26:00      2.75
##  4 536365    84029G    KNITTED UNION FLA~      6 2010-12-01 08:26:00      3.39
##  5 536365    84029E    RED WOOLLY HOTTIE~      6 2010-12-01 08:26:00      3.39
##  6 536365    22752     SET 7 BABUSHKA NE~      2 2010-12-01 08:26:00      7.65
##  7 536365    21730     GLASS STAR FROSTE~      6 2010-12-01 08:26:00      4.25
##  8 536366    22633     HAND WARMER UNION~      6 2010-12-01 08:28:00      1.85
##  9 536366    22632     HAND WARMER RED P~      6 2010-12-01 08:28:00      1.85
## 10 536367    84879     ASSORTED COLOUR B~     32 2010-12-01 08:34:00      1.69
## # i 541,899 more rows
## # i 2 more variables: CustomerID <dbl>, Country <chr>
```

2.Data prep

```
## [1] NA
```

```
##    InvoiceNo          StockCode          Description          Quantity
##  Length:541909      Length:541909      Length:541909      Min.   :-80995.00
##  Class :character   Class :character   Class :character   1st Qu.:      1.00
##  Mode  :character   Mode  :character   Mode  :character   Median :      3.00
##                                                           Mean   :      9.55
##                                                           3rd Qu.:     10.00
##                                                           Max.   : 80995.00
##
##   InvoiceDate                           UnitPrice          CustomerID
##  Min.   :2010-12-01 08:26:00.00   Min.   :-11062.06   Min.   :12346
##  1st Qu.:2011-03-28 11:34:00.00   1st Qu.:      1.25   1st Qu.:13953
##  Median :2011-07-19 17:17:00.00   Median :      2.08   Median :15152
##  Mean   :2011-07-04 13:34:57.16   Mean   :      4.61   Mean   :15288
##  3rd Qu.:2011-10-19 11:27:00.00   3rd Qu.:      4.13   3rd Qu.:16791
##  Max.   :2011-12-09 12:50:00.00   Max.   : 38970.00   Max.   :18287
##                                                       NA's   :135080
##    Country
##  Length:541909
##  Class :character
##  Mode  :character
##
##
##
##
```

3.Analysis: Present to CMO

3.1 CMO: Which item is the most sales in the UK?

```
## # A tibble: 3,849 x 2
##    Description                        sum
##    <chr>                            <dbl>
##  1 PAPER CRAFT , LITTLE BIRDIE      80998
##  2 MEDIUM CERAMIC TOP STORAGE JAR   76937
##  3 WORLD WAR 2 GLIDERS ASSTD DESIGNS 49197
##  4 JUMBO BAG RED RETROSPOT          42104
##  5 WHITE HANGING HEART T-LIGHT HOLDER 34768
##  6 ASSORTED COLOUR BIRD ORNAMENT    32757
##  7 POPCORN HOLDER                   29028
##  8 PACK OF 12 LONDON TISSUES        24343
##  9 BROCADE RING PURSE               22720
## 10 PACK OF 72 RETROSPOT CAKE CASES  22498
## # i 3,839 more rows
```

- ANS: PAPER CRAFT , LITTLE BIRDIE has the best seller in UK

3.2 CMO: Give me 5 items of the worst sales on Dec,2010?

```
## # A tibble: 5 x 2
##   Description                        sum
##   <chr>                            <dbl>
## 1 3D HEARTS  HONEYCOMB PAPER GARLAND    1
## 2 4 PINK DINNER CANDLE SILVER FLOCK     1
## 3 5 STRAND GLASS NECKLACE AMBER         1
```

```
## 4 AMBER FINE BEAD NECKLACE W TASSEL      1
## 5 ANT COPPER RED BOUDICCA BRACELET       1
```
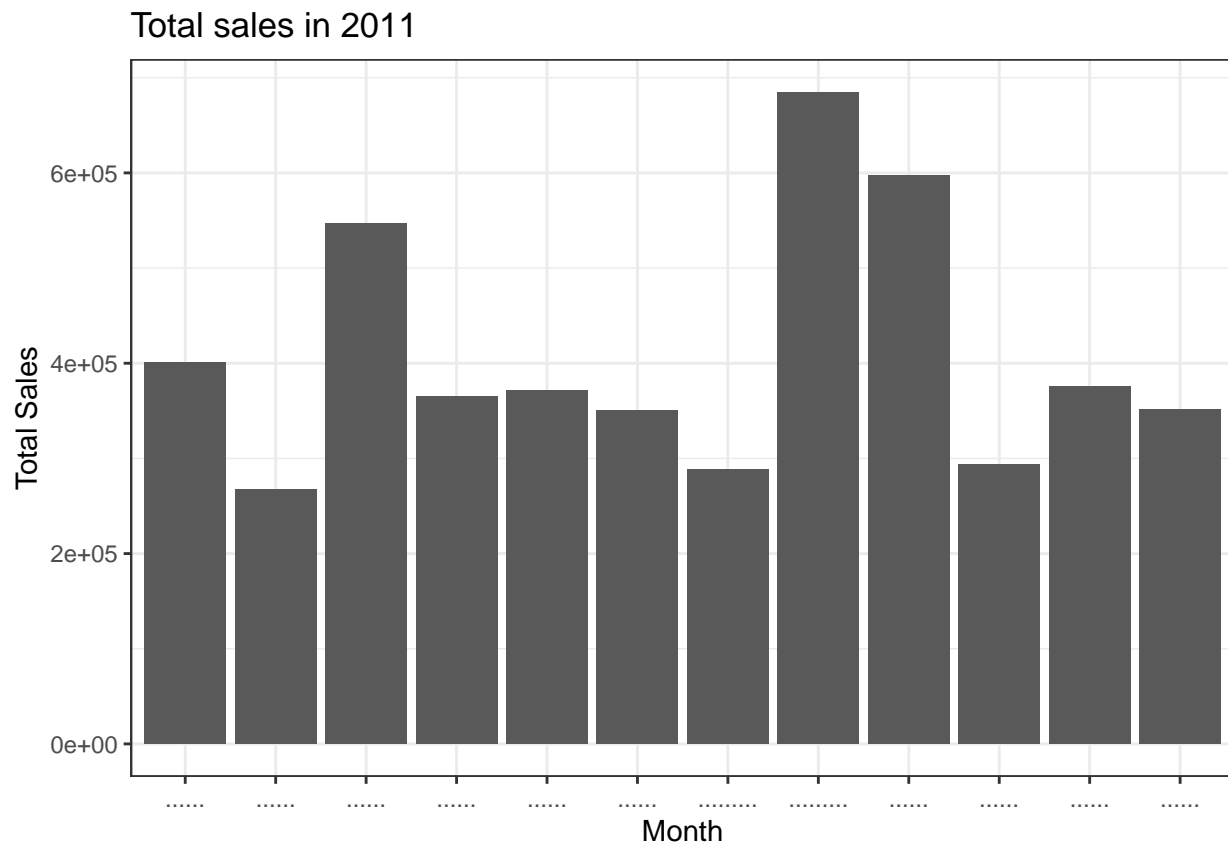
- ANS:
    1. 3D HEARTS HONEYCOMB PAPER GARLAND

    2. 5 STRAND GLASS NECKLACE AMBER

    3. 75 GREEN FAIRY CAKE CASES

    4. AMBER DROP EARRINGS W LONG BEADS

    5. AMBER FINE BEAD NECKLACE W TASSEL

3.3 CMO: How many transactions of PAPER CRAFT?

```
## # A tibble: 2 x 6
##   StockCode Description             Quantity InvoiceDate Country   InvoiceNo
##   <chr>     <chr>                      <dbl> <date>      <chr>     <chr>
## 1 23843     PAPER CRAFT , LITTLE BIRDIE   80995 2011-12-09  United K~ 581483
## 2 23843     PAPER CRAFT , LITTLE BIRDIE       3 2011-12-09  United K~ C581484
```

- ANS: 2 transactions

3.4 CMO: Please show me line chart of Total Sales in 2011

Total sales in 2011

4.Analysis: Present to CEO

4.1 CEO: Which month has the most revenue in 2011

```
## # A tibble: 12 x 2
##    month  Total_revenue
##    <chr>          <dbl>
## 1            460202.
## 2            505543.
## 3          526969.
## 4            588259.
## 5            618446.
## 6            650620.
## 7            673152.
## 8            748997.
## 9            793294.
## 10           973772.
## 11          1101534.
## 12         1217847.
```

- ANS: December has the most revenue in 2011

4.2 CEO: Which quarter has the fewest revenue in 2011

```
## # A tibble: 4 x 2
##   Quarter Total_revenue
##   <chr>           <dbl>
## 1 Q1          1666907.
## 2 Q2          2047835.
## 3 Q3          2297543.
## 4 Q4          2846350.
```

- ANS: Q1 has the fewest revenue in 2011

4.3 CEO: Which country is the best choice to expand project

```
## # A tibble: 37 x 2
##    Country              Total_revenue
##    <chr>                        <dbl>
## 1 Saudi Arabia                  155.
## 2 Bahrain                       548.
## 3 Czech Republic                856.
## 4 RSA                          1004.
## 5 Brazil                       1144.
## 6 European Community           1313
## 7 Lithuania                    1661.
## 8 Lebanon                      1694.
## 9 United Arab Emirates         1902.
## 10 Unspecified                 2667.
## # i 27 more rows
```

- ANS: UAE has the most revenue country. So we might plan to expand project in there.

4.4 CEO: Describe revenue in each country and month by visualization in UK and UAE

```
## `summarise()` has grouped output by 'Country'. You can override using the
## `.groups` argument.
```


Total Revenue in 2011