

Project Data Analysis for Finance: Performa Cabang

Author : Imam Aris Munandar

```
#load Data loan disbursement as df_loan
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
#load data from csv
df_loan <- read.csv('https://dqlab-dataset.s3-ap-southeast-1.amazonaws.com/loan_disbursement.csv', stringsAsFactors = F)
dplyr::glimpse(df_loan)
```

```
## Observations: 9,754
## Variables: 5
## $ loan_id      <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15...
## $ tanggal_cair <chr> "2020-01-06", "2020-01-06", "2020-01-06", "2020-0...
## $ cabang       <chr> "AA", "AA", "AA", "AA", "AA", "AA", "AB", "AB", "...
## $ agen        <chr> "AA-1", "AA-1", "AA-1", "AA-2", "AA-2", "AA-2", "...
## $ amount      <int> 320000, 440000, 200000, 430000, 360000, 220000, 3...
```

```
# Memfilter data bulan Mei 2020, dan jumlahkan data per cabang
library(scales)
df_loan_mei <- df_loan %>%
  filter(tanggal_cair >= '2020-05-01', tanggal_cair <= '2020-05-31') %>%
  group_by(cabang) %>%
  summarise(total_amount = sum(amount))
df_loan_mei
```

```
## # A tibble: 22 x 2
##   cabang total_amount
##   <chr>      <int>
## 1 AA      75710000
## 2 AB      81440000
## 3 AC      83990000
## 4 AD      76080000
## 5 AE      54200000
## 6 AF      68040000
## 7 AG      74080000
## 8 AH      73840000
## 9 AI      46640000
## 10 AJ      43580000
## # ... with 12 more rows
```

```
#data 5 cabang dengan total amount paling besar
```

```
df_loan_mei %>%
  arrange(desc(total_amount)) %>%
  mutate(total_amount = comma(total_amount)) %>%
  head(5)
```

```
## # A tibble: 5 x 2
##   cabang total_amount
##   <chr>   <chr>
## 1 AC      83,990,000
## 2 AB      81,440,000
## 3 AD      76,080,000
## 4 AA      75,710,000
## 5 AG      74,080,000
```

```
#data 5 cabang dengan total amount paling kecil
df_loan_mei %>%
  arrange(total_amount) %>%
  mutate(total_amount = comma(total_amount)) %>%
  head(5)
```

```
## # A tibble: 5 x 2
##   cabang total_amount
##   <chr>   <chr>
## 1 AV      30,280,000
## 2 AS      31,740,000
## 3 AT      34,840,000
## 4 AU      35,610,000
## 5 AO      39,120,000
```

```
#Menghitung umur cabang (dalam bulan)

df_cabang_umur <- df_loan %>%
  group_by(cabang) %>%
  summarise(pertama_cair = min(tanggal_cair)) %>%
  mutate(umur = as.numeric(as.Date('2020-05-15') - as.Date(pertama_cair)) %/% 30)
df_cabang_umur
```

```
## # A tibble: 22 x 3
##   cabang pertama_cair umur
##   <chr>   <chr>       <dbl>
## 1 AA      2020-01-06         4
## 2 AB      2020-01-06         4
## 3 AC      2020-01-06         4
## 4 AD      2020-01-06         4
## 5 AE      2020-02-03         3
## 6 AF      2020-02-03         3
## 7 AG      2020-02-03         3
## 8 AH      2020-02-03         3
## 9 AI      2020-03-02         2
## 10 AJ      2020-03-02         2
## # ... with 12 more rows
```

Gabungkan data umur dan performa mei Selanjutnya untuk membandingkan data umur dan performa di bulan mei, terlebih dahulu perlu digabungkan dulu data-data yang sudah dibuat sebelumnya dengan menggunakan fungsi `inner_join`, lalu simpan sebagai `df_loan_mei_umur`.

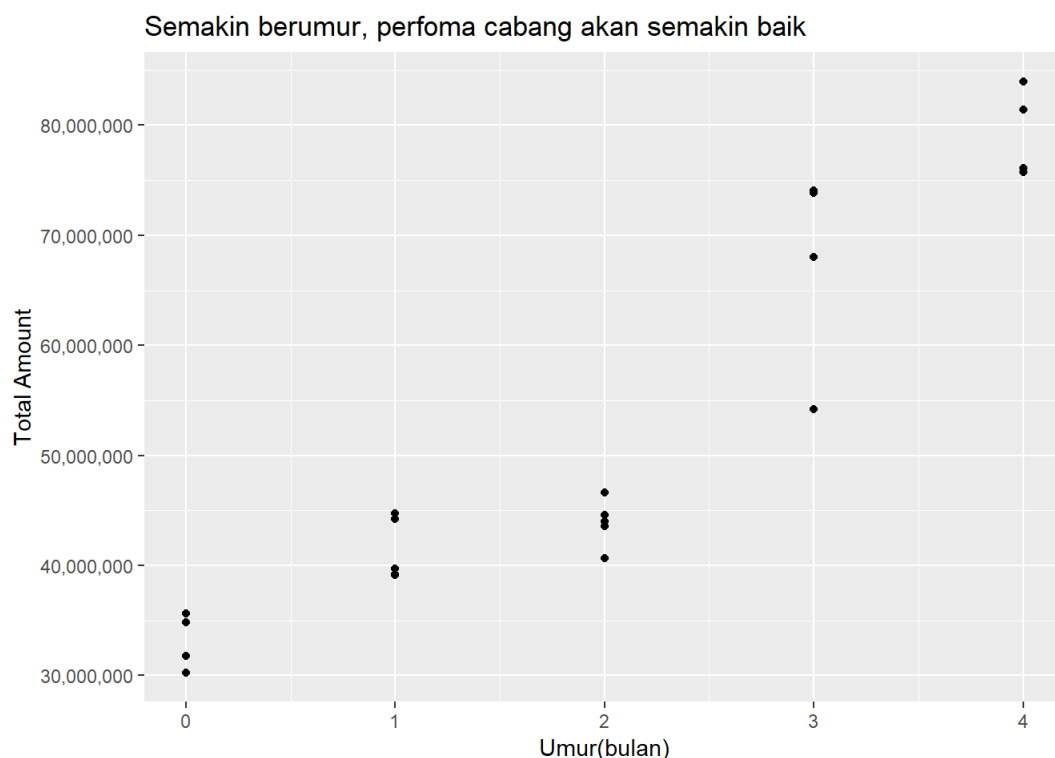
```
df_loan_mei_umur <- df_loan %>%
  inner_join(df_cabang_umur, by = 'cabang')
df_loan_mei_umur
```

```
## # A tibble: 22 x 4
##   cabang pertama_cair umur total_amount
##   <chr>   <chr>      <dbl>      <int>
## 1 AA     2020-01-06      4    75710000
## 2 AB     2020-01-06      4    81440000
## 3 AC     2020-01-06      4    83990000
## 4 AD     2020-01-06      4    76080000
## 5 AE     2020-02-03      3    54200000
## 6 AF     2020-02-03      3    68040000
## 7 AG     2020-02-03      3    74080000
## 8 AH     2020-02-03      3    73840000
## 9 AI     2020-03-02      2    46640000
## 10 AJ    2020-03-02      2    43580000
## # ... with 12 more rows
```

```
library(ggplot2)
```

```
## Registered S3 methods overwritten by 'ggplot2':
##   method      from
## [.quosures   rlang
## c.quosures   rlang
## print.quosures rlang
```

```
#Plot relasi umur dan performa mei
ggplot(df_loan_mei_umur, aes(x = umur, y = total_amount)) +
  geom_point() +
  scale_y_continuous(labels = scales::comma) +
  labs(title = "Semakin berumur, perfoma cabang akan semakin baik",
       x = "Umur(bulan)",
       y = "Total Amount")
```



Kesimpulan

Terlihat bahwa ada pola semakin tua cabang, maka performa nya semakin baik. Hal ini karena cabang tersebut masih berkembang sehingga belum sampai pada performa maksimal.

Akan tetapi pada masing - masing umur itu juga ada cabang yang performanya dibawah yang lain.

Selanjutnya akan dianalisis cabang yang performanya lebih rendah dari yang lain pada umur yang sama

Mencari cabang yang perfoma rendah untuk setiap umur

```
library(dplyr)
library(scales)
df_loan_mei_flag <- df_loan_mei_umur %>%
  group_by(umur) %>%
  mutate(Q1 = quantile(total_amount, 0.25),
         Q3 = quantile(total_amount, 0.75),
         IQR = (Q3 - Q1)) %>%
  mutate(flag = ifelse(total_amount < (Q1 - IQR), 'rendah', 'baik'))

df_loan_mei_flag %>%
  filter(flag == 'rendah') %>%
  mutate_if(is.numeric, funs(comma))
```

```
## `mutate_if()` ignored the following grouping variables:
## Column `umur`
```

```
## Warning: funs() is soft deprecated as of dplyr 0.8.0
## please use list() instead
##
## # Before:
## funs(name = f(.))
##
## # After:
## list(name = ~f(.))
## This warning is displayed once per session.
```

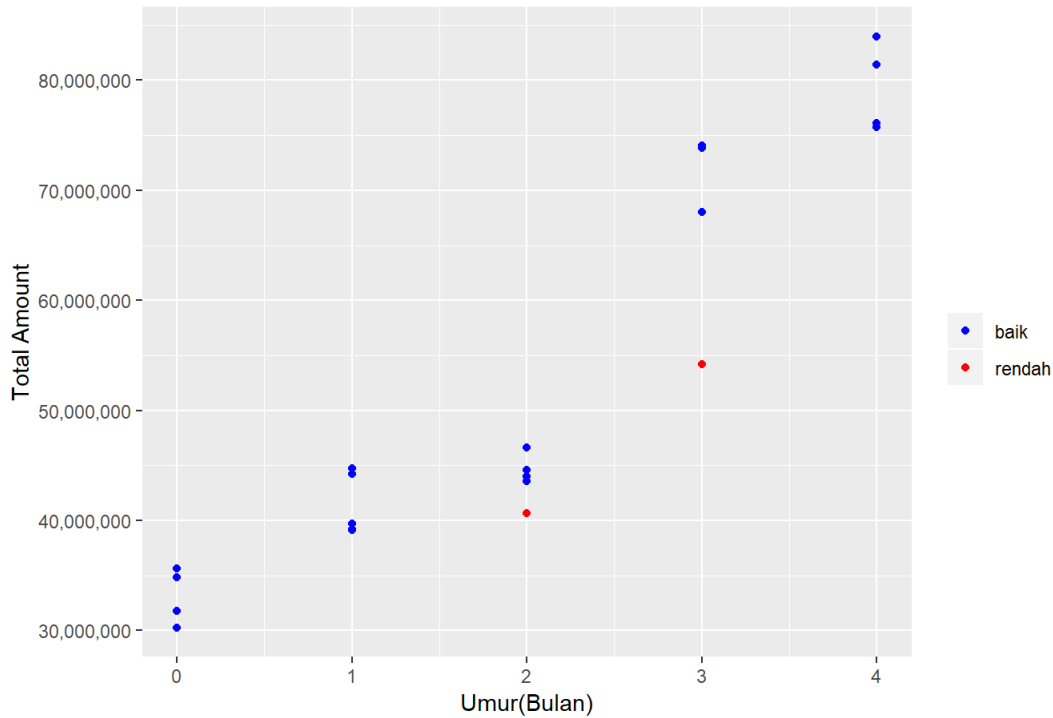
```
## # A tibble: 2 x 8
## # Groups:   umur [2]
##   cabang pertama_cair umur total_amount Q1      Q3      IQR      flag
##   <chr> <chr>      <dbl> <chr>    <chr>    <chr>    <chr>
## 1 AE    2020-02-03      3 54,200,000 64,580,0~ 73,900,0~ 9,320,0~ rend~
## 2 AL    2020-03-02      2 40,650,000 43,580,0~ 44,590,0~ 1,010,0~ rend~
```

Color Change Scatterplot

```
library(ggplot2)

ggplot(df_loan_mei_flag, aes(x = umur, y = total_amount)) +
  geom_point(aes(color = flag)) +
  scale_color_manual(breaks = c("baik", "rendah"),
                    values=c("blue", "red")) +
  scale_y_continuous(labels = scales::comma) +
  labs(title = "cabang berpeforma rendah padahal tidak termasuk bottom 5 nasional",
       color = "",
       x = "Umur (Bulan)",
       y = "Total Amount")
```

cabang berperforma rendah padahal tidak termasuk bottom 5 nasional



perbandingan performa cabang di umur yang sama

```
df_loan_mei_flag %>%
  filter(umur == 3) %>%
  inner_join(df_loan, by = 'cabang') %>%
  filter(tanggal_cair >= '2020-05-01', tanggal_cair <= '2020-05-31') %>%
  group_by(cabang, flag) %>%
  summarise(jumlah_hari = n_distinct(tanggal_cair),
            agen_aktif = n_distinct(agen),
            total_loan_cair = n_distinct(loan_id),
            avg_amount = mean(amount),
            total_amount = sum(amount)) %>%
  arrange(total_amount) %>%
  mutate_if(is.numeric, funs(comma))
```

```
## `mutate_if()` ignored the following grouping variables:
## Column `cabang`
```

```
## # A tibble: 4 x 7
## # Groups:   cabang [4]
##   cabang flag jumlah_hari agen_aktif total_loan_cair avg_amount
##   <chr> <chr> <chr>      <chr>      <chr>      <chr>
## 1 AE    rend~ 21         3        175        309,714
## 2 AF    baik  21         3        225        302,400
## 3 AH    baik  21         3        241        306,390
## 4 AG    baik  21         3        241        307,386
## # ... with 1 more variable: total_amount <chr>
```

perbandingan performa agen pada cabang yang rendah

```
df_loan_mei_flag %>%
  filter(umur == 3, flag == 'rendah') %>%
  inner_join(df_loan, by = 'cabang') %>%
  filter(tanggal_cair >= '2020-05-01', tanggal_cair <= '2020-05-31') %>%
  group_by(cabang, agen) %>%
  summarise(jumlah_hari = n_distinct(tanggal_cair),
            total_loan_cair = n_distinct(loan_id),
            avg_amount = mean(amount),
            total_amount = sum(amount)) %>%
  arrange(total_amount) %>%
  mutate_if(is.numeric, funs(comma))
```

```
## `mutate_if()` ignored the following grouping variables:
## Column `cabang`
```

```
## # A tibble: 3 x 6
## # Groups:   cabang [1]
##   cabang agen jumlah_hari total_loan_cair avg_amount total_amount
##   <chr> <chr> <chr>         <chr>         <chr>         <chr>
## 1 AE    AE-3    4             16             310,625       4,970,000
## 2 AE    AE-2    18            73             320,274       23,380,000
## 3 AE    AE-1    21            86             300,581       25,850,000
```

Lihat perbandingan performa agen pada cabang yang paling baik umur 3 bulan

Pada tabel sebelumnya, terlihat pula bahwa ada cabang yang punya 3 agen, tapi performa nya jauh diatas cabang AE, bahkan yang paling tinggi diantara cabang lain pada umur tersebut, lebih tinggi dari yang mempunya 4 agen cabang tersebut adalah cabang AH.

```
df_loan %>%
  filter(cabang == 'AH') %>%
  filter(tanggal_cair >= '2020-05-01', tanggal_cair <= '2020-05-31') %>%
  group_by(cabang, agen) %>%
  summarise(jumlah_hari = n_distinct(tanggal_cair),
            total_loan_cair = n_distinct(loan_id),
            avg_amount = mean(amount),
            total_amount = sum(amount)) %>%
  arrange(total_amount) %>%
  mutate_if(is.numeric, funs(comma))
```

```
## `mutate_if()` ignored the following grouping variables:
## Column `cabang`
```

```
## # A tibble: 3 x 6
## # Groups:   cabang [1]
##   cabang agen jumlah_hari total_loan_cair avg_amount total_amount
##   <chr> <chr> <chr>         <chr>         <chr>         <chr>
## 1 AH    AH-3    19             74             303,649       22,470,000
## 2 AH    AH-1    21             81             301,358       24,410,000
## 3 AH    AH-2    21             86             313,488       26,960,000
```

Kesimpulan

Berdasarkan analisis tersebut, dapat disimpulkan bahwa rendahnya performa dari cabang AE adalah karena salah satu agen yang melakukan pencairan hanya 4 hari dalam 1 bulan, padahal agen lain bisa aktif 21 hari.

Hal ini membuat total amount dari agen tersebut hanya 20% dibandingkan agen yang lainnya. Sedangkan pada cabang AH, performanya sangat baik karena ketiga agen melakukan pencairan hampir / selalu setiap hari kerja. 2 orang full 21 hari 1 orang 19 hari. Sehingga performa nya terjaga dengan baik. Selanjutnya perlu dianalisis lebih lanjut kenapa ada agen yang hanya aktif beberapa hari saja dalam sebulan. Untuk kedepannya setiap agen agar dipastikan untuk bisa aktif setiap hari nya, sehingga bisa menjaga performa cabang.