Data-540-Lab-3

Dhun Sheth

2023-10-27

Part 1

```
con <- dbConnect(RMySQL::MySQL(), user='dsheth', password='32376881',
dbname='tpch', host='cosc304.ok.ubc.ca')
```

Part 2

```
part_2_query <- "
select ps_partkey, ps_suppkey, ps_availqty, year(l_shipdate) as year,
sum(l_quantity) as shippedQuantity
from partsupp join lineitem on l_partkey = ps_partkey and l_suppkey = ps_suppkey
group by ps_partkey, ps_suppkey, ps_availqty, year(l_shipdate)
order by ps_partkey, ps_suppkey, year(l_shipdate)
"
part_2_res <- dbGetQuery(con, part_2_query)
```

```
## Warning in .local(conn, statement, ...): Decimal MySQL column 4 imported as
## numeric
```

```
part_2_res[1:10,]
```

```
##    ps_partkey ps_suppkey ps_availqty year shippedQuantity
## 1           1          2       3325 2015              50
## 2           1          2       3325 2018               5
## 3           1         27       8076 2013              62
## 4           1         27       8076 2015              43
## 5           1         27       8076 2016              34
## 6           1         27       8076 2018             134
## 7           1         52       3956 2015               4
## 8           1         52       3956 2017              40
## 9           1         52       3956 2018              49
## 10          1         77       4069 2013              76
```

Part 3,4

```
con_sqlite <- dbConnect(RSQLite::SQLite(), ":memory:")


dbWriteTable(con_sqlite, "part_3_sqlite", part_2_res)
part_3_4_data <- dbGetQuery(con_sqlite,"SELECT * FROM part_3_sqlite")
part_3_4_data[1:10,]
```

```
##    ps_partkey ps_suppkey ps_availqty year shippedQuantity
## 1           1          2        3325 2015              50
## 2           1          2        3325 2018               5
## 3           1         27        8076 2013              62
## 4           1         27        8076 2015              43
## 5           1         27        8076 2016              34
## 6           1         27        8076 2018             134
## 7           1         52        3956 2015               4
## 8           1         52        3956 2017              40
## 9           1         52        3956 2018              49
## 10          1         77        4069 2013              76
```

Part 5

```
part_5_query <- "
select *
from part_3_sqlite
where year = 2018 and ps_availqty < shippedQuantity
limit 5
"

part_5_data <- dbGetQuery(con_sqlite, part_5_query)

print('Products with insufficient inventory based on 2018 sales:')
```

```
## [1] "Products with insufficient inventory based on 2018 sales:"
```

```
part_5_data
```

```
##   ps_partkey ps_suppkey ps_availqty year shippedQuantity
## 1         50         76          43 2018              67
## 2         51          2         138 2018             189
## 3         81         57          58 2018              68
## 4        217         18          20 2018              92
## 5        281         63          51 2018             114
```

Part 6

```
part_6_query <- "
select year, shippedQuantity
from part_3_sqlite
```

```
where ps_partkey = 217 and ps_suppkey = 18
"

part_6_data <- dbGetQuery(con_sqlite, part_6_query)
part_6_reg <- lm(shippedQuantity ~ year, data=part_6_data)
summary(part_6_reg)
```

```
##
## Call:
## lm(formula = shippedQuantity ~ year, data = part_6_data)
##
## Residuals:
##       1       2       3       4       5       6
## -29.571  34.457  26.486 -33.486  -2.457   4.571
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -13980.914  15066.238  -0.928    0.406
## year             6.971      7.475   0.933    0.404
##
## Residual standard error: 31.27 on 4 degrees of freedom
## Multiple R-squared:  0.1786, Adjusted R-squared:  -0.02674
## F-statistic: 0.8698 on 1 and 4 DF,  p-value: 0.4038
```

```
reg_est_2019 <- part_6_reg$coefficients[2] * 2019 + part_6_reg$coefficients[1]
print('Estimated Sales for 2019 based on the regression is:')
```

```
## [1] "Estimated Sales for 2019 based on the regression is:"
```

```
reg_est_2019[['year']]
```

```
## [1] 94.4
```

```
# plot(part_6_data)
```

Part 7

```
part_7_query <- "
select *, round((ps_availqty/shippedQuantity),2) as percentageOverstocked
from part_3_sqlite
where year = 2018 and shippedQuantity >= 6
order by percentageOverstocked desc
limit 10
"
part_7_data <- dbGetQuery(con_sqlite, part_7_query)
part_7_data
```

```
##    ps_partkey ps_suppkey ps_availqty year shippedQuantity percentageOverstocked
## 1        1828         58        9958 2018               6               1659.67
```

```
## 2         200         53        9408 2018                6               1568.00
## 3         287         15        9210 2018                6               1535.00
## 4        1695         19        9154 2018                6               1525.67
## 5        1388         65        8718 2018                6               1453.00
## 6         631         32        8673 2018                6               1445.50
## 7        1457         75        8526 2018                6               1421.00
## 8        1891         78        9897 2018                7               1413.86
## 9        1523         24        8461 2018                6               1410.17
## 10        834         68        9559 2018                7               1365.57
```

```
dbDisconnect(con)
```

```
## [1] TRUE
```

```
dbDisconnect(con_sqlite)
```