

Regression Model for Study Cafe Revenue :

Which one impacts the Study Cafe Revenue most among Space, # of School and Floor.

Name: Jincheol Park

Due date: November 3rd, 2024

Problem Motivation

The growth in the popularity of study cafes has transformed these spaces into crucial environments for students and professionals looking for a conducive place to focus and work. Study cafes offer unique amenities that attract customers, but their profitability and revenue can vary significantly based on specific attributes. Understanding which factors most significantly affect study cafe revenue is essential to optimizing operational efficiency, making strategic improvements, and ensuring sustainable profitability.

In this analysis, I aim to investigate the impact of four key factors—Space (square footage of the Study Cafe), Number of Schools in the area, and Floor (building level of the Study Cafe) and Price (per 2hour)—on the revenue of study cafes. These factors were chosen based on their direct or indirect influence on customer convenience, accessibility, and demand potential.

Research Objective

The primary objective of this research is to analyze and quantify the impact of Space (total area of the Study Cafe), Number of Schools in the area, Floor (building level of the Study Cafe) and Price on the revenue generated by study cafes. By employing a regression model, this study aims to determine which of these factors has the most substantial influence on study cafe revenue, providing actionable insights for study cafe owners, investors, and other stakeholders in the industry.

The specific research objectives are as follows:

1. Evaluate the Influence of Space on Revenue:

To understand the relationship between the size of the study cafe and its revenue. This objective seeks to determine if larger spaces correlate with higher revenue, or if revenue generation is more closely tied to other factors, independent of the physical space.

2. Assess the Impact of Nearby Schools on Revenue:

To examine whether the proximity and number of educational institutions within a specific radius positively affect study cafe revenue. This objective aims to determine if areas with more schools have a higher demand for study spaces, leading to increased profitability.

3. Analyze the Effect of Floor Location on Revenue:

To investigate how the cafe's floor level within a building affects its revenue. This will provide insights into whether more accessible, lower-floor locations drive more customer traffic and higher revenue compared to cafes on upper floors.

4. Price and Customer Perception:

Pricing is a crucial factor that directly impacts customer decisions and revenue. Study cafes that offer competitive pricing may attract more budget-conscious students, while premium pricing might appeal to customers seeking a quieter or more exclusive atmosphere. Understanding the balance between price and revenue generation will provide insights into optimal pricing strategies.

By achieving these objectives, this study will clarify which of the variables—Space, Number of Schools, Floor and Price—has the greatest impact on revenue, thereby guiding better decision-making for study cafe site selection, spatial design, and overall investment strategy.

Literature review

The study of factors impacting revenue in service-based businesses, such as study cafes, draws on insights from multiple disciplines, including real estate economics, urban planning, and consumer behavior. While the specific literature on study cafes is limited, relevant findings from studies on coworking spaces, coffee shops, and other similar spaces can provide valuable context. This literature review explores three critical factors—Space, Number of Schools, Floor and Price—that influence revenue in similar settings, setting the foundation for this study's investigation.

1. Space and Revenue Potential

The size of a retail or service space often influences its revenue capacity, with larger spaces typically accommodating more customers and offering greater revenue potential. Studies on urban green spaces have found a positive relationship between space and revenue, primarily due to increased seating capacity and additional amenities (¹*Institutional factors affecting urban green space provision – from a local government revenue perspective*).

In the context of study cafes, similar principles may apply. Larger spaces can host more customers, creating a conducive environment for studying and working. Yet, if space is underutilized or poorly designed, it may not effectively increase revenue. This study builds on these findings by examining whether larger study cafe spaces directly correlate with higher revenue and identifying the extent of this influence.

2. Proximity to Schools

Proximity to educational institutions has consistently been identified as a factor in catering to students, including coffee shops, bookstores, and study spaces. Numerous studies indicate that houses located near schools attract significant faculty, thus boosting revenue (²*House prices and school choice: Evidence from Chicago's magnet schools' proximity lottery*).

¹ Institutional factors affecting urban green space provision - from a local government revenue perspective : <https://www.tandfonline.com/doi/abs/10.1080/09640568.2018.1541231>

² House prices and school choice: Evidence from Chicago's magnet schools' proximity lottery : <https://onlinelibrary.wiley.com/doi/abs/10.1111/jors.12447>

Similarly, the presence of multiple schools within a certain radius can signify a concentrated target market, increasing the likelihood of steady patronage. While this factor has been well-documented in broader retail contexts, there is limited specific research on its impact on study cafes. This study seeks to bridge this gap by investigating whether the number of nearby schools is a significant predictor of revenue for study cafes, providing insights for optimal location selection.

3. Floor Level and Accessibility

The accessibility of a location, often determined by its floor level, has been shown to affect customer convenience and, consequently, revenue. Ground-level businesses are generally more accessible and visible, factors that have been linked to higher customer traffic and revenue in various service industries (³*Does Better Physical Accessibility Lead to Higher Sales Revenues? The Case of Slovenian Restaurant SMEs*). However, upper-floor locations may pose accessibility challenges, reducing customer willingness to visit unless there is a strong, specific motivation.

4. Pricing Strategy and Customer Demand

Pricing is a crucial factor for service-based businesses, as it directly affects customer perception and willingness to pay. A study on relationship between Customer demand(⁴*The relationship between customer value and pricing strategies: an empirical test*) indicates that price-sensitive customers, such as students, respond positively to competitive pricing structures that offer perceived value without compromising quality.

Data and method Explanation

1. Space and Customer Capacity (Independent Variable):

- Larger spaces can generally accommodate more customers, which could intuitively lead to higher revenue. However, space utilization efficiency can vary; merely increasing the area might not guarantee increased revenue if it does not align with customer demand or comfort. Understanding the quantitative relationship between space and revenue is key to making more informed decisions about space allocation and lease costs.

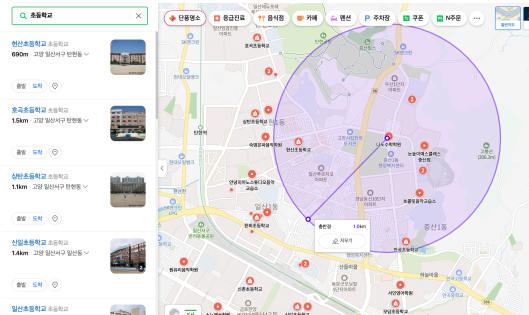
2. Number of Schools (Independent Variable):

- The proximity to educational institutions often drives foot traffic in study cafes, as students are one of the primary customer bases. Areas with a higher number of nearby schools could exhibit higher demand for study spaces. By quantifying this relationship, cafe owners can better select locations and determine the extent to which the surrounding educational landscape influences revenue.

³ Does Better Physical Accessibility Lead to Higher Sales Revenues? The Case of Slovenian Restaurant SMEs : <https://academica.turistica.si/index.php/AT-TIJ/article/view/607>

⁴ The relationship between customer value and pricing strategies: an empirical test : <https://www.emerald.com/insight/content/doi/10.1108/10610421211276321/full/html>

- Number of Schools data (Elementary School/Middle School/High School) is collected by using ⁵'Naver' Map Radius function. This data is counted within 1km radius as below picture.



3. Floor and Accessibility (Independent Variable):

- The floor level of a cafe in a building could impact customer convenience and accessibility. Higher floors may be less attractive to customers due to accessibility constraints, while ground-level locations might attract more foot traffic. This analysis aims to explore whether the floor level significantly influences revenue.
- Number of Floor data is collected by using 'Naver' Map Address information.

4. Price (Independent Variable):

- Although Price range have the variety, price by 2 hour is set on this data set.

5. Revenue (Dependent Variable):

- Revenue Data is collected by using ⁶'Open up' tracking function. Data is set by (Min+Max)/2



| Revenue | Space | Price | Floor | Elementary | Middle | High | Name |
|---------|-------|-------|-------|------------|--------|------|-------------------|
| 350 | 120 | 3000 | 4 | 5 | 3 | 3 | 작심스터디카페 일산하늘마을점 |
| 700 | 140 | 4000 | 2 | 5 | 3 | 3 | 공본스터디카페 하늘마을센터 |
| 1600 | 200 | 3000 | 2 | 4 | 2 | 2 | 이루지오 스터디카페 |
| 600 | 140 | 3000 | 4 | 4 | 2 | 2 | 공간너머스터디카페 |
| 1050 | 160 | 2500 | -1 | 7 | 2 | 2 | 자술 스터디카페 |
| 800 | 160 | 3000 | 4 | 7 | 2 | 2 | 플랜에이스터디카페 후곡센터 |
| 1250 | 160 | 4000 | 5 | 7 | 2 | 2 | 공부인스터디카페 일산후곡센터 |
| 650 | 140 | 3000 | 5 | 7 | 2 | 2 | 모모스터디카페 |
| 500 | 120 | 3000 | 2 | 8 | 2 | 2 | 디플레이스 스터디카페 일산주업점 |
| 500 | 140 | 3000 | 5 | 8 | 2 | 2 | 블루문 스터디카페 주업점 |
| 1000 | 160 | 3000 | 9 | 4 | 1 | 0 | 브릴리스터디카페 독서실 마두역점 |
| 300 | 120 | 3000 | 5 | 5 | 1 | 0 | 르하임스터디카페 일산마두점 |
| 1800 | 200 | 3000 | 5 | 3 | 2 | 1 | 플랜에이스터디카페 일산백마센터 |
| 750 | 160 | 4000 | 3 | 3 | 2 | 1 | 작심스터디카페 일산백마점 |
| 700 | 160 | 3000 | 5 | 3 | 2 | 1 | 위년스터디카페 일산백마회원가점 |
| 300 | 130 | 4000 | 4 | 4 | 2 | 2 | 공부인스터디카페 풍동센터 |
| 350 | 120 | 3000 | 4 | 4 | 2 | 2 | 르하임스터디카페 일산풍동점 |
| 350 | 140 | 4000 | 4 | 4 | 2 | 2 | 초심스터디카페 일산풍동점 |
| 350 | 120 | 3000 | 5 | 3 | 4 | 2 | 아카데미라운지 중산독서실 |
| 450 | 140 | 4000 | 4 | 2 | 3 | 1 | 위년스터디카페 일산중산점 |

⁵ Naver Map : <https://map.naver.com/p/search/스터디카페/place/12006627?c=15.38,0,0,0,dh>

⁶ Open up : <https://www.openub.com/>

1. Regression Model

Regression model is used to check which impact Revenue most among independent variable.

Normalized value from independent value is used to improve regression model efficiency.

```
normalize <- function(x) {return ((x - min(x)) / (max(x) - min(x)))}
data <- Ilsan_Studay_Cafe[, -c(1,8)]
data_normalized <- as.data.frame(lapply(data, normalize))
data_normalized <- cbind(data_normalized, C = Ilsan_Studay_Cafe[1])

> md <- lm(Revenue ~ Space + Price + Floor + Elementary + Middle + High, data = data_normalized)
> summary(md)
```

Call:

```
lm(formula = Revenue ~ Space + Price + Floor + Elementary + Middle +
    High, data = data_normalized)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|--------|
| -238.81 | -68.57 | 21.86 | 86.72 | 260.51 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|--------------|
| (Intercept) | 101.16 | 240.34 | 0.421 | 0.681 |
| Space | 1400.82 | 140.45 | 9.974 | 1.85e-07 *** |
| Price | -84.62 | 124.61 | -0.679 | 0.509 |
| Floor | 70.87 | 238.57 | 0.297 | 0.771 |
| Elementary | 227.83 | 179.67 | 1.268 | 0.227 |
| Middle | 182.49 | 269.37 | 0.677 | 0.510 |
| High | -31.01 | 247.68 | -0.125 | 0.902 |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 167.3 on 13 degrees of freedom

Multiple R-squared: 0.8962, Adjusted R-squared: 0.8483

F-statistic: 18.7 on 6 and 13 DF, p-value: 1.065e-05

As the result of summary for Regression model, independent value except for Space variable has 0.05 higher p-value. That means we can only trust the Space variable on this model. And Estimate Std for Space variable is highest. And this value have the lot of difference with other independent variable. Hence, we can predict Space variable have the highest impact on the Revenue.

2. Partial F-test

Partial F-test is used for improving Regression model trust efficiency. As the result of 3x direction, AIC value of backward and both have lower value(203.81) than forward(210.18). Step function of backward and both direction only make the Space and Elementary variable exist in the Regression model.

#Partial F-test

```
forward_md = step(md,scope=list(lower=~1,upper=md),direction="forward")
backward_md = step(md,direction="backward")
both_md = step(md,scope=list(upper=md),direction="both")
```

Backward direction model same as both direction only have Space and Elementary variable. P-value of Elementary variable is 0.1 and it means it is hard to trust. Therefore, This model mean only Space variable highly impact Revenue.

```
> summary(backward_md)
```

Call:

```
lm(formula = Revenue ~ Space + Elementary, data = data_normalized)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|--------|-------|--------|
| -227.92 | -114.29 | 25.97 | 95.42 | 229.86 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|--------------|
| (Intercept) | 168.8 | 79.8 | 2.116 | 0.0494 * |
| Space | 1371.2 | 118.5 | 11.573 | 1.75e-09 *** |
| Elementary | 198.8 | 115.2 | 1.726 | 0.1026 |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 152.4 on 17 degrees of freedom

Multiple R-squared: 0.8874, Adjusted R-squared: 0.8741

F-statistic: 66.97 on 2 and 17 DF, p-value: 8.69e-09

3. VIF

VIF method is used for checking Multicollinearity of this full Regression model. As all of VIF value is lower than 5, all of independent variable have no Multicollinearity.

```
> vif(md)
```

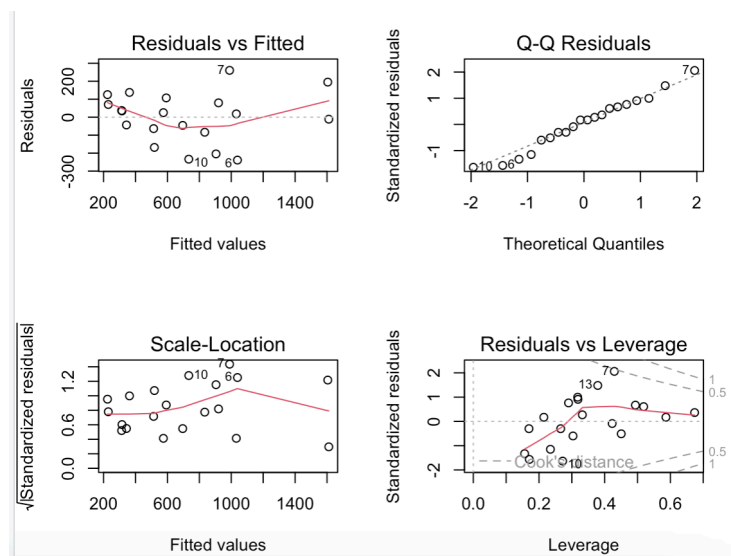
| Space | Price | Floor | Elementary | Middle | High |
|----------|----------|----------|------------|----------|----------|
| 1.194836 | 1.167648 | 1.422809 | 2.067049 | 2.461802 | 2.969643 |

4. Residuals Analysis

Residuals vs Fitted means Residuals have linearity or not. As red line have no linearity, residual expected not to have linearity.

Q-Q Residuals can expect Residuals have norm distribution or not. As most of dot is not highly out of dotted line, residual expected to have norm distribution.

Scale-Location can expect Residuals have homogeneity or not. As the red line is close to horizontal line, residual expected to have homogeneity



The Shapiro-Wilks test for the residuals was carried out to check the normality of the residuals. The following is the result of it. From the output, the p-value of residuals $0.7384 > 0.05$ (significance level) implies the distribution of the residual is not significantly different from the normal distribution.

```
> shapiro.test(res)

Shapiro-Wilk normality test

data:  res
W = 0.96922, p-value = 0.7384
```

The graph below is the result of the studentized Breusch-Pagan test for checking heteroscedasticity. The obtained p-value of 0.1517 is way higher than 0.05. Therefore, the heteroscedasticity exists among residuals.

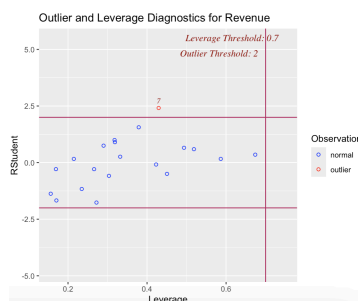
```
> bptest(md)

studentized Breusch-Pagan test

data:  md
BP = 9.4124, df = 6, p-value = 0.1517
```

5. Outlier Analysis

Ols_plot_resid_lev is used for detecting outlier. As the below picture, #7 index is detected as outliers. But, method of not excepting #7 index for model study is to be used because R_Student value is not that high and only having 20x dataset for model study.



Analysis Result

As the Space variable is highly impact to Revenue on this model, it is expected that the higher Space the Study Cafe have the higher Revenue is. Other variable of # of school, Floor, Price also seems important. But, there only have 20x dataset. It follows to have higher value of P-value on those variable. Hence, other variable also seems to be trusted if dataset is enough.

of seats in Study Cafe is also to be important independent variable. But, it can be correlated with Space variable. So, when using this variable, correlation need to be considered.

Year of establishment is also to be important independent variable. It is expected that the more Study Cafe is established recently, the higher Revenue is.

of people by region is also to be important independent variable. It is expected the more people by region, the higher Revenue on this region is. But, dataset have to be enough to consider this variable on this model.

In the conclusion, Space have to be considered firstly when establishing Study Cafe. Although Price/Floor/# of School is important variable to be considered when establishing Study Cafe, Space is more important variable than others.

Reference List

- Institutional factors affecting urban green space provision – from a local government revenue perspective : <https://www.tandfonline.com/doi/abs/10.1080/09640568.2018.1541231>
- House prices and school choice: Evidence from Chicago's magnet schools' proximity lottery : <https://onlinelibrary.wiley.com/doi/abs/10.1111/jors.12447>
- Does Better Physical Accessibility Lead to Higher Sales Revenues? The Case of Slovenian Restaurant SMEs : <https://academica.turistica.si/index.php/AT-TIJ/article/view/607>
- The relationship between customer value and pricing strategies: an empirical test : <https://www.emerald.com/insight/content/doi/10.1108/10610421211276321/full/html>
- Naver Map : <https://map.naver.com/p/search/스터디카페/place/12006627?c=15.38,0,0,0,dh>
- Open up : <https://www.openub.com/>