

Empirical Research

Note Taking Application

Group-19

Aditya Rajesh Patil	180101004
Kartikeya Saxena	180101034
Milind Prabhu	180101091
Kushal Sangwan	180101096

What is Empirical Research?

The process of collecting and analyzing data of user behaviour in a controlled experiment system is known as empirical research. It is not just limited to study of human behaviour only, any study involving observation collection and analysis for its purpose is empirical research.



Research Question Formulation

Broad Question: How does a user's aesthetic judgement depend on the default note interface?

Why did we choose this research question?

Many users who use note taking applications rarely bother to change the default parameters like font size, font face, color etc. so it is crucially important that our default interface has the aesthetics which satisfy the user and prevent the irritating task of changing the settings every time they login.

For others who do bother, it might happen that the app default is not the best option out there. And every time they open a new note, they have to change the default settings to something better. In an attempt to answer this question, we aim to provide a neat default view that won't need much altering each time an user opens a new note.

Since this is a very broad question, it is quite untestable, so it is better to break down the question into multiple testable questions so it is easy to test and also generalize well.



Research Question Formulation

Testable Questions

Q1: How does aesthetic score (in a scale of 1-10) depends on font size(3), font type(2) of the default note interface.

Q2: How does aesthetic score (in a scale of 1-10) depends on background theme(3), cursor style(2) of the default note interface.

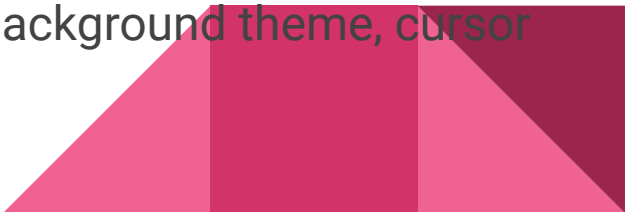


Research Question Formulation

Null Hypothesis

1. Aesthetic score (in a scale of 1-10) does not depend on the font size, font type of the note interface.
2. Aesthetic score (in a scale of 1-10) does not depend on the background theme, cursor style of the note interface.

Alternate Hypothesis

1. Aesthetic score (in a scale of 1-10) depends on the font size, font type of the note interface.
 2. Aesthetic score (in a scale of 1-10) depends on the background theme, cursor style of the note interface.
- 

Identification of Variables

Independent Variables:

Font size, font type will be our independent variables (factors) for our first testable question whereas background theme and cursor style will be the factors for the second testable question.


Font size is an ordinal variable which can take any positive integer as its value whereas font face, background theme and cursor style will be nominal variables.

To determine the set of levels these factors(independent variables) can take, we adopted pilot study.



Identification of Variables

Pilot Study: We went across various popular note taking applications, text editors and IDE like Microsoft OneNote, Google Docs, VS Code, Atom, VIM and came to the following conclusive levels.

1. **Font Style :-** We went through Google Docs(Arial style) , Microsoft Word(Calibri style) and OneNote and found out that each application uses a unique font style as their default . Hence we decided to pick up a new one for our app and took two levels one each from Serif and Sans Serif types.
 2. **Font Size :-** We found that while Google Docs uses font size of 11 pts , VS Code uses size of 17 pts. Hence we took the average(14 pts) and made out three levels , one equal to it and others below and above it.
 3. **Cursor Style :-** Most note-taking apps use blinking cursors in the shape of a line. But some editors like VIM use a stagnant block cursor. Hence we chose these two levels for the cursor style.
 4. **Background Theme :-** Many applications allow multiple background themes for their users but the default one is sometimes light and sometimes dark. Hence we took light, dark 2 and dark 1(which is in between light and dark 2) .
- 


Identification of Variables

FACTORS	LEVELS		
Font Style(S)	Lora(Serif)		Nota Sans(Sans Serif)
Font Size(N)	11	14	17

FACTORS	LEVELS		
Cursor Style(C)	Line		Block
Background Theme (T)	Light	Dark1	Dark2

Identification of Variables

Dependent Variables:

1. **Rating:** The participants rate the default note interface (more specifically its attributes or factors) on an interval scale from 1 to 10 with 1 indicating that the participants do not like the aesthetics and 10 indicating that they like the aesthetics. The interval scale has the property that the difference between any two consecutive ratings is the same. For example the difference between an interface with rating 4 and interface with rating 3 is the same as the difference between an interface with rating 7 and an interface with rating 6.
- 

Identification of Variables

Control Variables:

1. **User Gender:** A student can be either female or male. And as the concept of aesthetic appearance may be different for these two, we decided to fix the gender to be male as the relation between our dependent variable(rating) and user gender is of no interest to our study
2. **User Program:** Students of Bachelors,Masters or Phd may rate the interfaces differently . Hence we fixed these variable to Bachelors program.
3. **Time of experiment:** A user's perception may differ when he is given little time to rate a interface as compared to when he is given more. Hence we fixed it to 10 seconds



Identification of Variables

Confounding Variables:

1. **Practice Effect:** It happens when all users are presented the interfaces in a particular order enforcing experience and expectations with each trial . To tackle this the latin square method was followed.



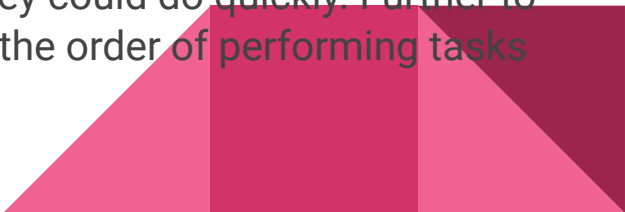
Design of Experiment

Experimental Setup: Each participant was asked to perform 12 tasks (6 tasks for each of the two testable questions). The participant profile, their numbers and the order of the assigned tasks were three key things to be decided about the way of conducting the experiment.

Participant Profile: All participants were male college students enrolled in an undergraduate program.

Number of participants: Twelve participants rated all 6 interfaces for each of the two testable question.

Assigning Tasks: The tasks were assigned to the participants in within-subject manner. This means that each participant has to perform just 12 tasks which they could do quickly. Further to avoid practice effect, latin rectangle method was used for deciding the order of performing tasks for participants.



Design of Experiment

P#1	R1	R2	R3	R4	R5	R6
P#2	R2	R3	R4	R5	R6	R1
P#3	R3	R4	R5	R6	R1	R2
P#4	R4	R5	R6	R1	R2	R3
P#5	R5	R6	R1	R2	R3	R4
P#6	R6	R1	R2	R3	R4	R5
P#7	R1	R2	R3	R4	R5	R6
P#8	R2	R3	R4	R5	R6	R1
P#9	R3	R4	R5	R6	R1	R2
P#10	R4	R5	R6	R1	R2	R3
P#11	R5	R6	R1	R2	R3	R4
P#12	R6	R1	R2	R3	R4	R5

Latin Rectangle: Any interface is seen at the i-th position by an equal number of participants. This is done to decrease the practice effect.

Latin Square Entry	Testable Research Ques 1	Testable Research Ques 2
R1	Lora, 11	Light, Line
R2	Lora, 14	Dark1, Line
R3	Lora, 17	Dark2, Line
R4	Nota, 11	Light, Block
R5	Nota, 14	Dark1, Block
R6	Nota, 17	Dark2, Block



Interfaces for Testable Research Question 1

Search

Q

Semester 6 Notebook

CS345

CS568

Frequent Itemset Mining

ECLAT Algorithm

Apriori Algorithm

CS331

CS501

CS348

Semester 5 Notebook

ECLAT Algorithm

LOGOUT

FILE

EDIT

VIEW

INSERT

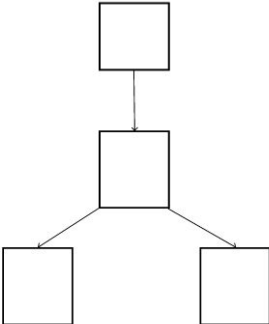
FORMAT

Itemset Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but itemset also the leap in, itemset to electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing itemset Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of itemset Ipsum

```
#include<iostream>

using namespace std;

int main() {
    int n;
    cout << "Enter a number.\n";
    cin >> n;
    cout << "Your number is: " << n << "\n";
    return 0;
}
```



Interface# 1

<Lora, 11pts>

Search

Q

Semester 6 Notebook

CS345

CS568

Frequent Itemset Mining

ECLAT Algorithm

Apriori Algorithm

CS331

CS501

CS348

Semester 5 Notebook

ECLAT Algorithm

LOGOUT

FILE

EDIT

VIEW

INSERT

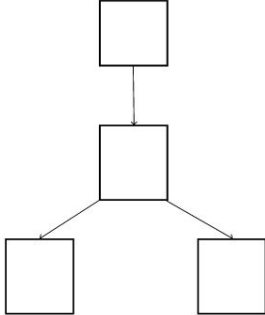
FORMAT

Itemset Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but itemset also the leap in, itemset to electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing itemset Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of itemset Ipsum

```
#include<iostream>

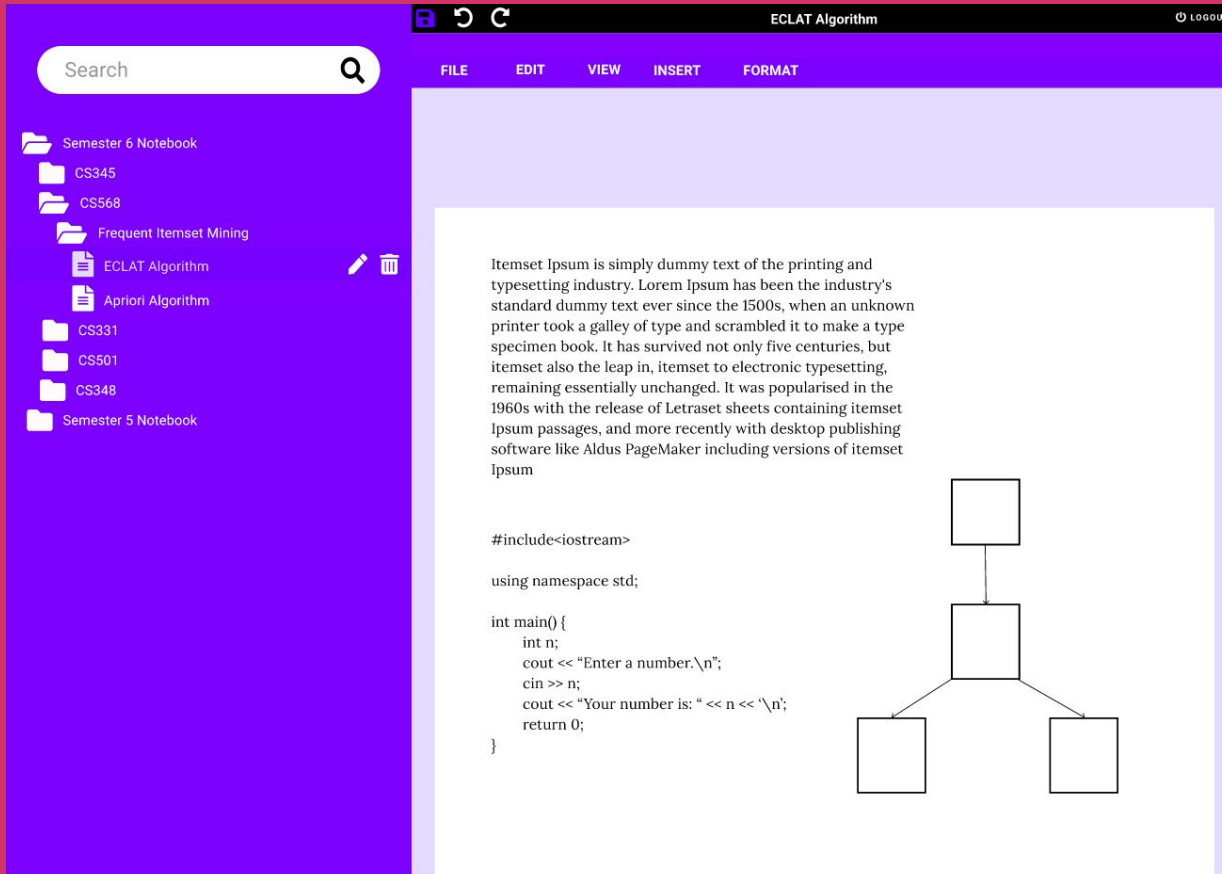
using namespace std;

int main() {
    int n;
    cout << "Enter a number.\n";
    cin >> n;
    cout << "Your number is: " << n << "\n";
    return 0;
}
```



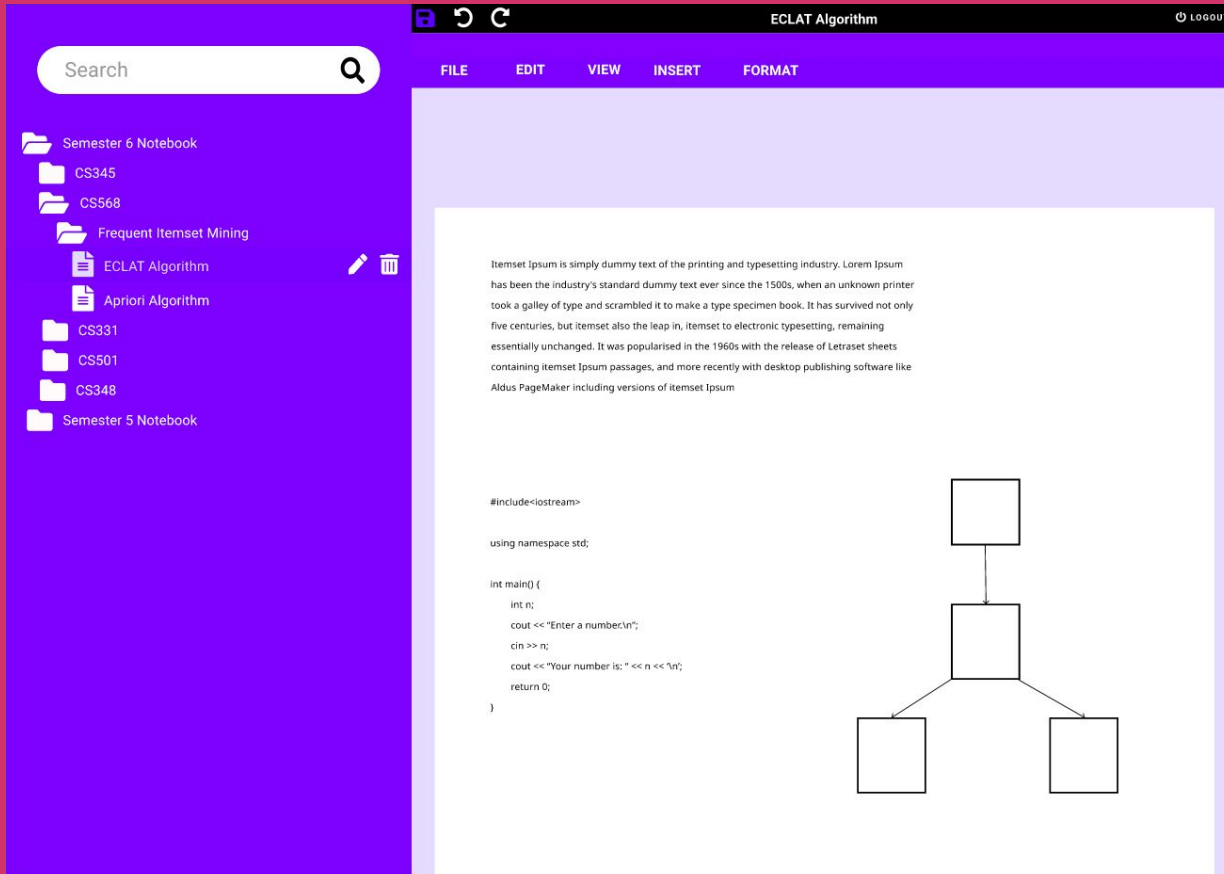
Interface# 2

<Lora, 14pts>



Interface# 3

<Lora, 17pts>



Interface# 4

<Noto Sans, 11pts>

Search

Q

Semester 6 Notebook

CS345

CS568

Frequent Itemset Mining

ECLAT Algorithm

Apriori Algorithm

CS331

CS501

CS348

Semester 5 Notebook

ECLAT Algorithm

LOGOUT

FILE

EDIT

VIEW

INSERT

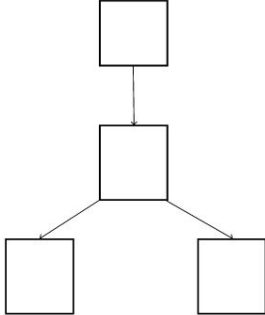
FORMAT

Itemset Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but itemset also the leap in, itemset to electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing itemset Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of itemset Ipsum

```
#include<iostream>

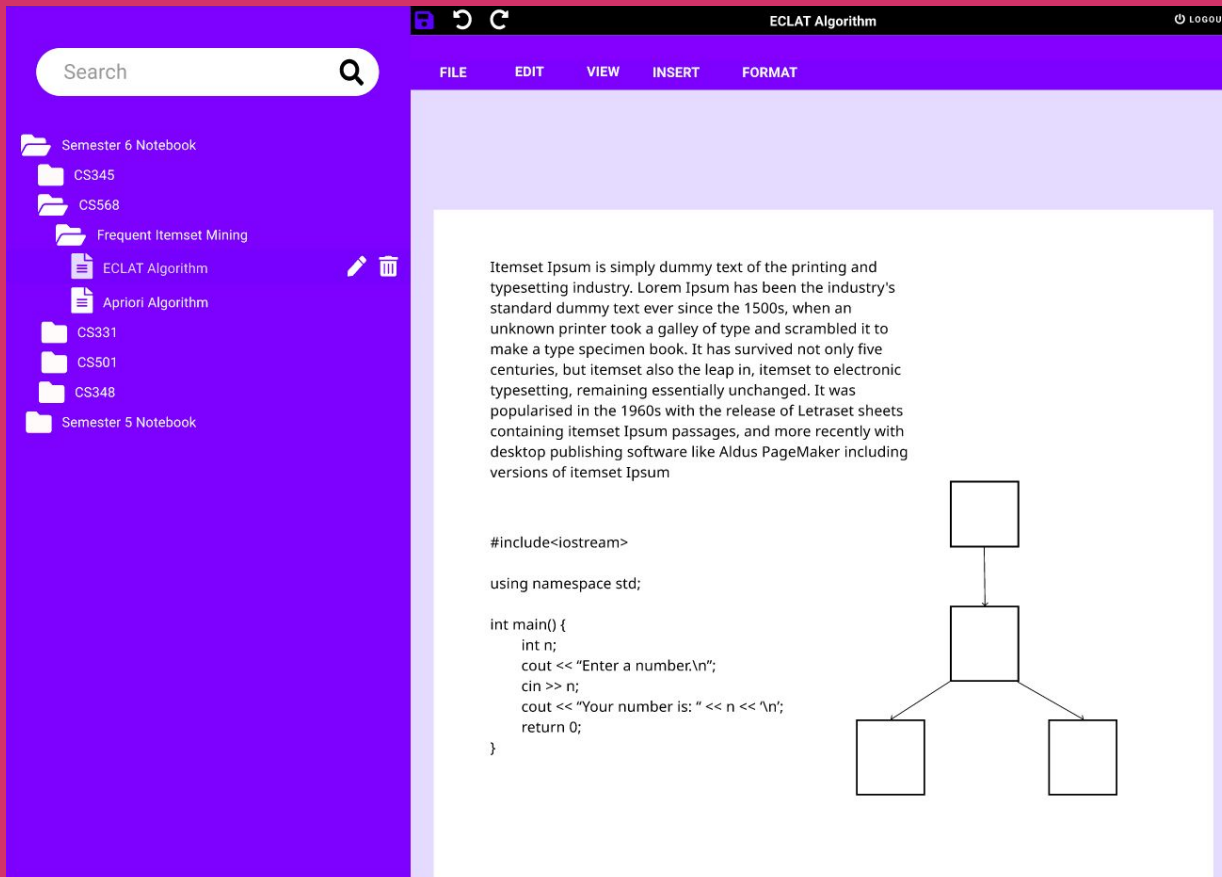
using namespace std;

int main() {
    int n;
    cout << "Enter a number.\n";
    cin >> n;
    cout << "Your number is: " << n << "\n";
    return 0;
}
```



Interface# 5

<Noto Sans, 14pts>

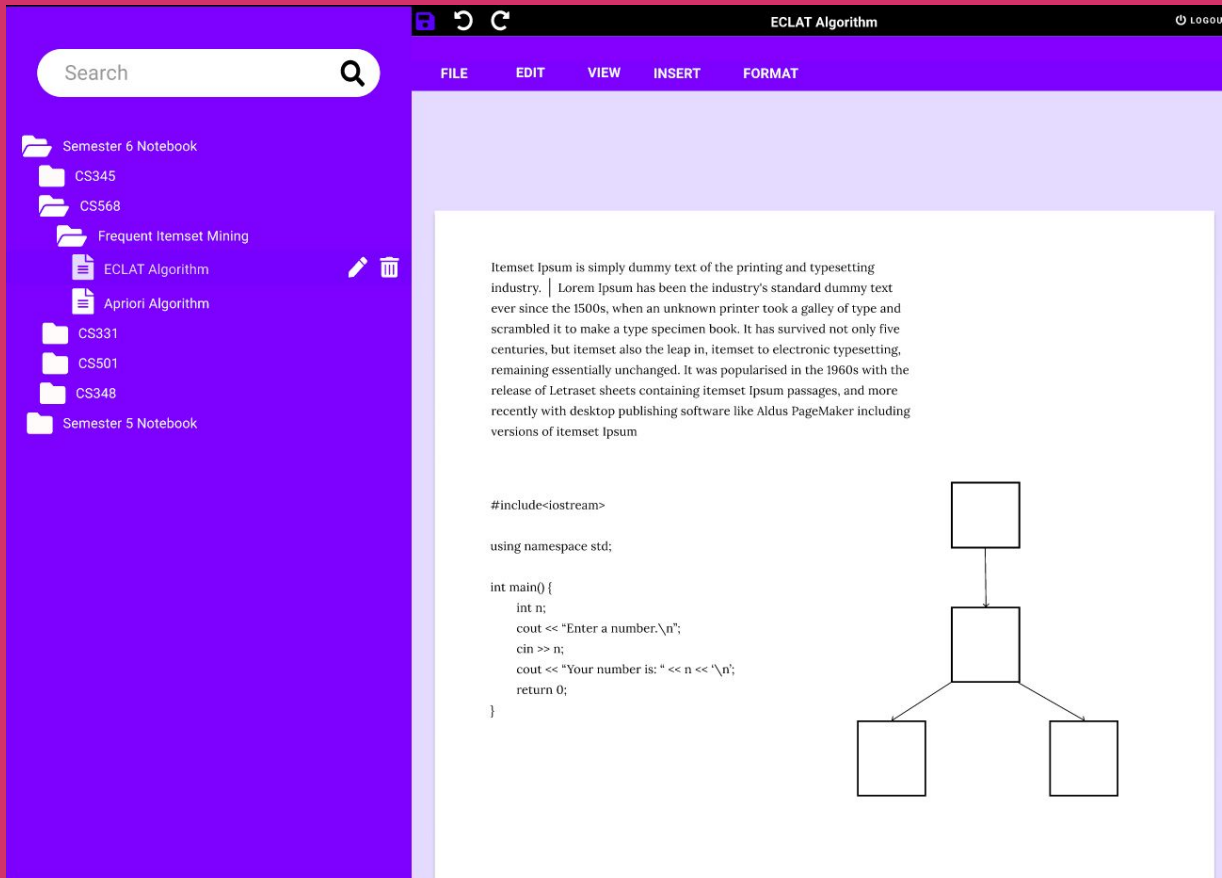


Interface# 6

<Noto Sans, 17pts>

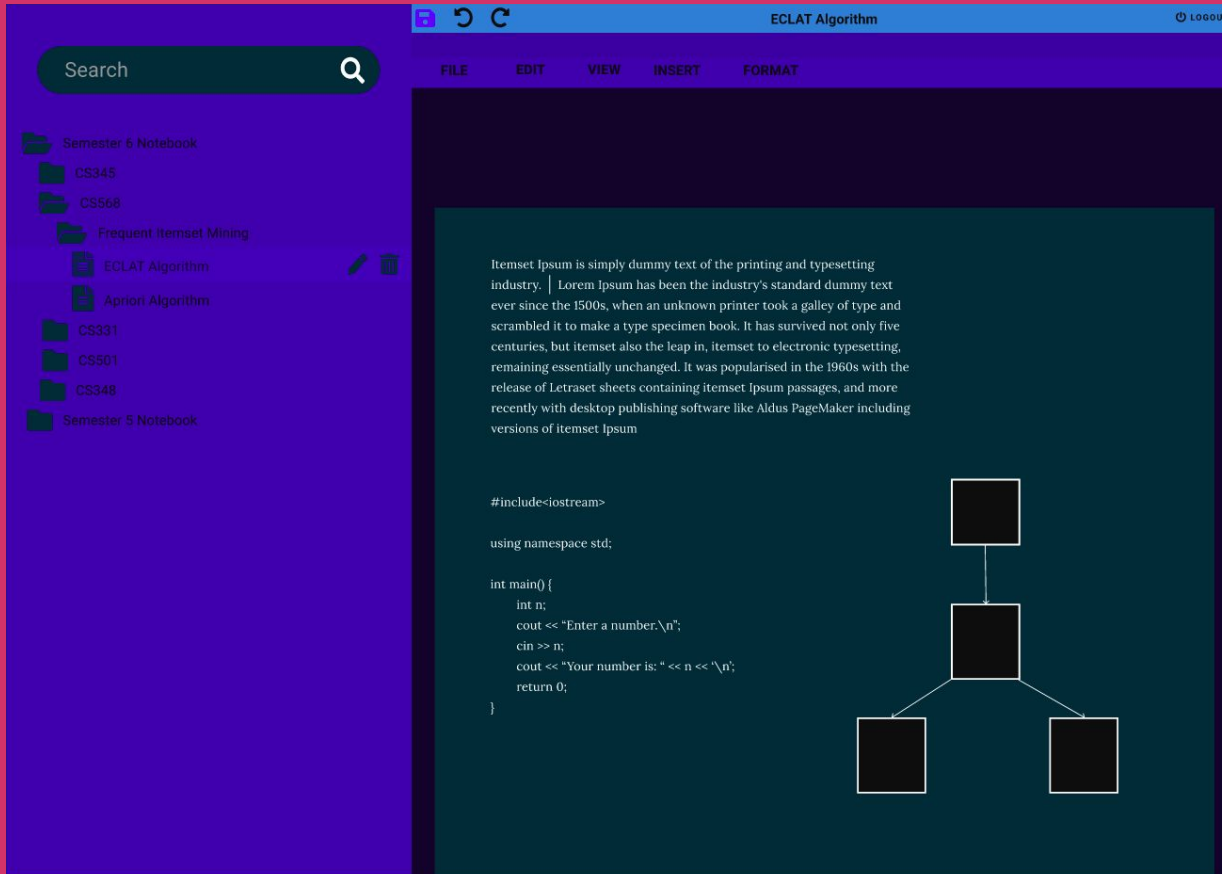


Interfaces for Testable Research Question 2



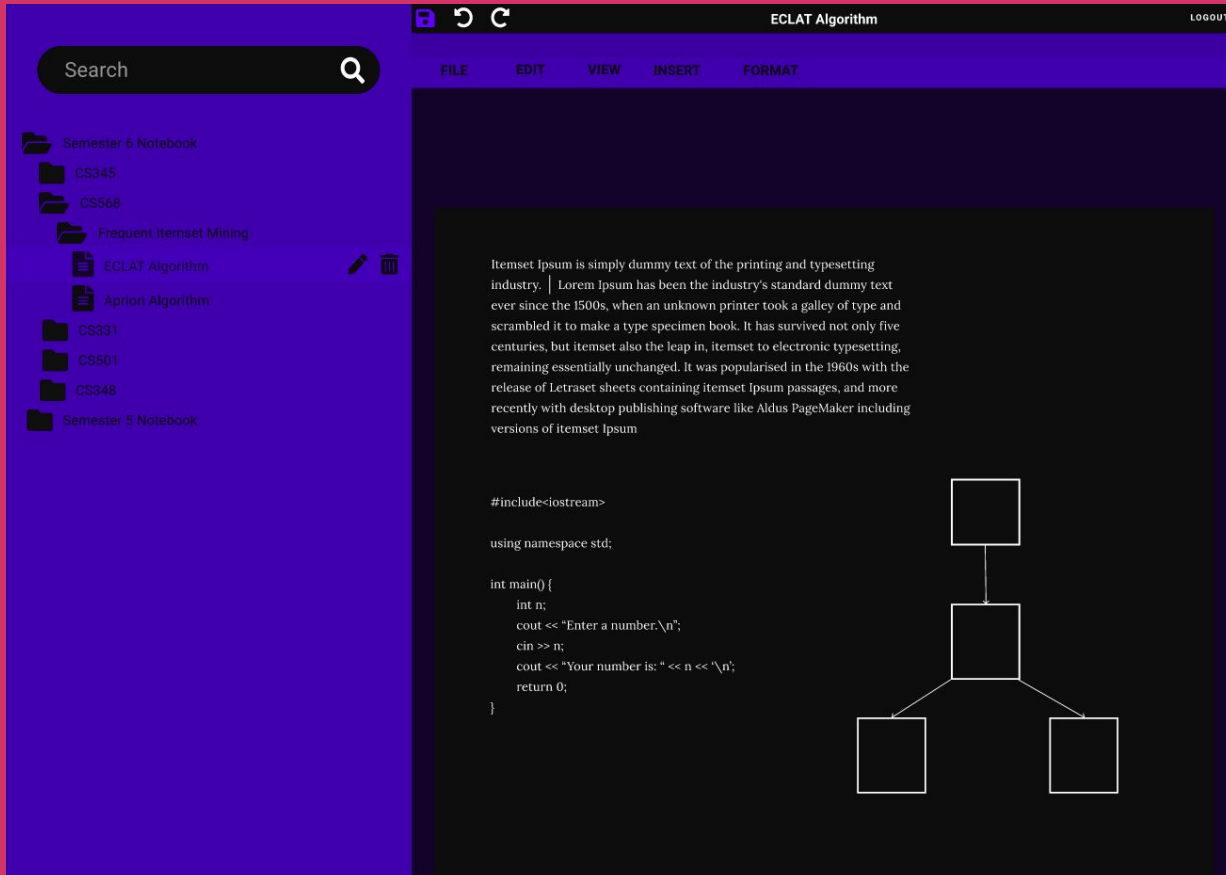
Interface# 1

<Light Theme, Line Cursor>



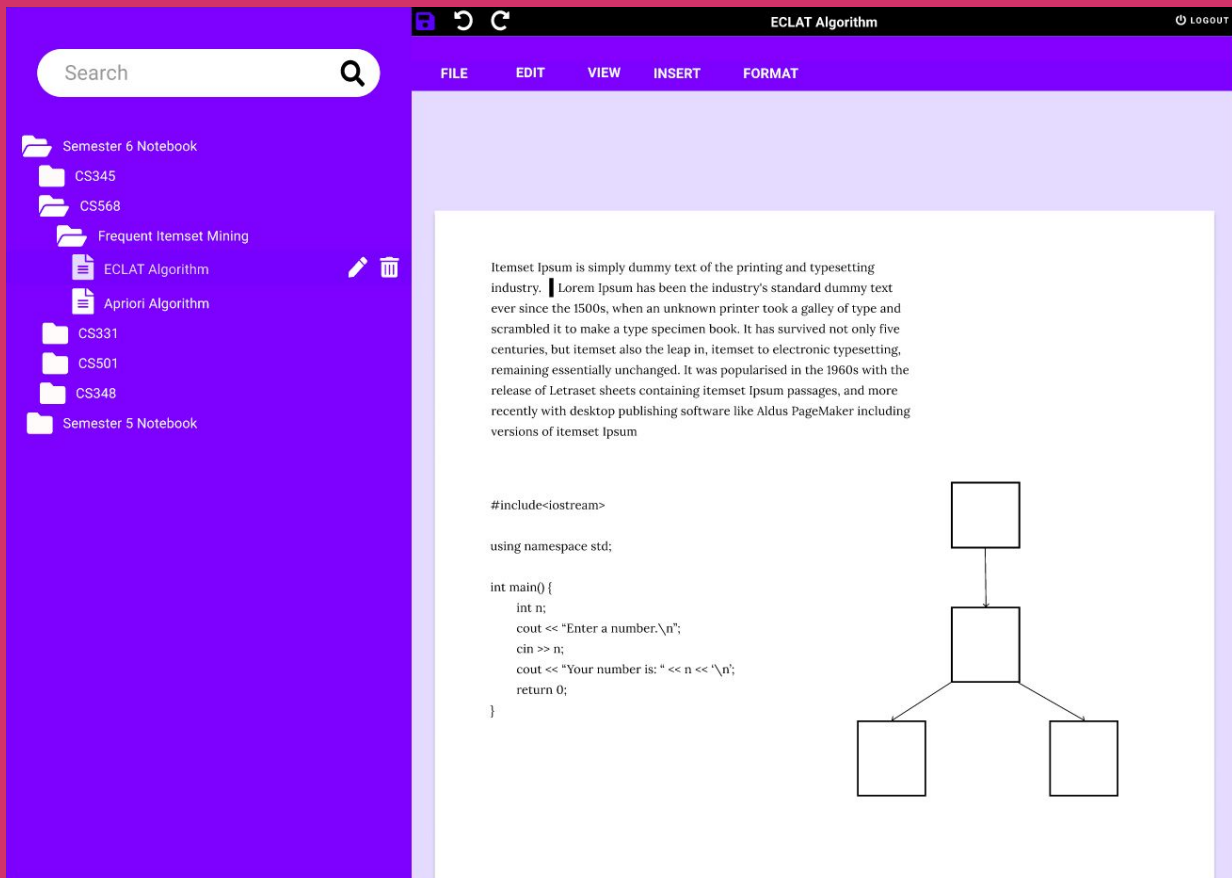
Interface# 2

<Dark 1 Theme, Line Cursor>



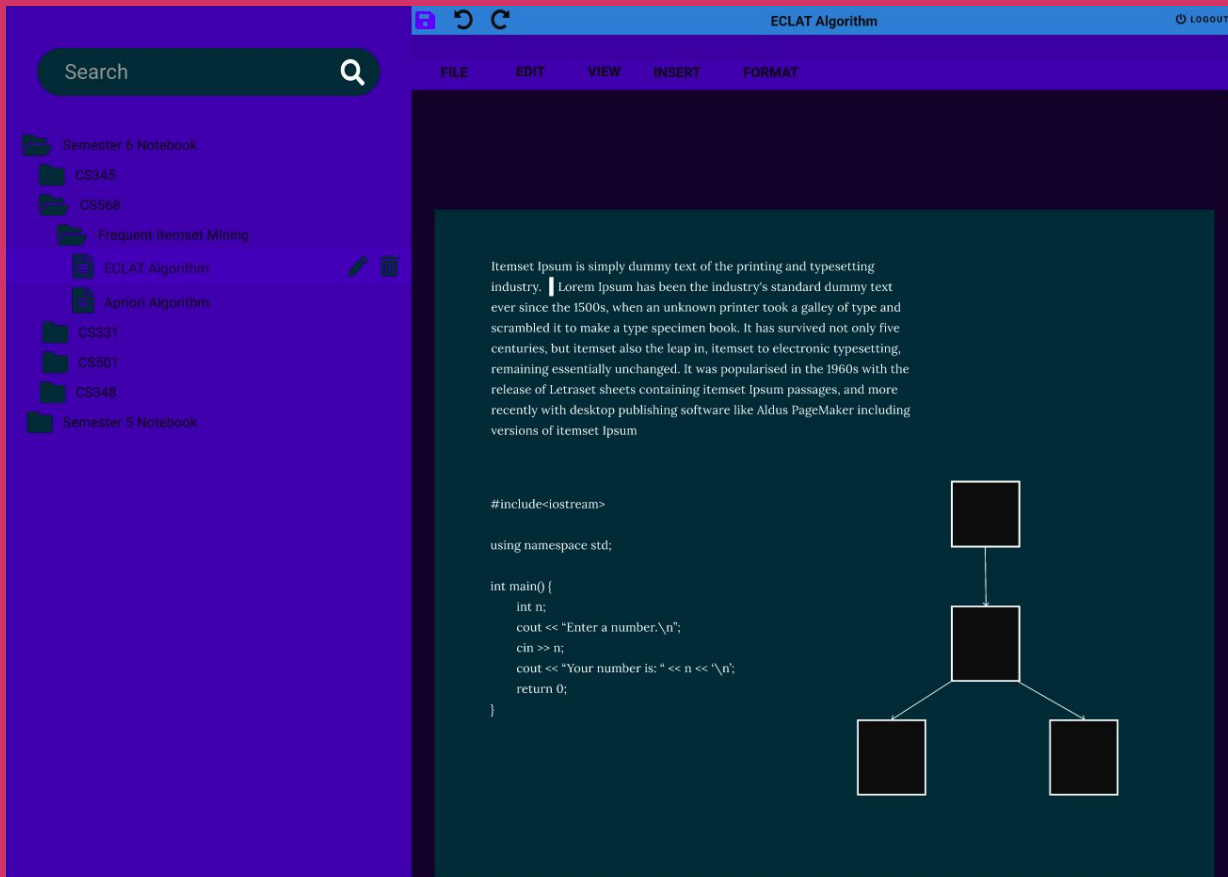
Interface# 3

<Dark 2 Theme, Line Cursor>



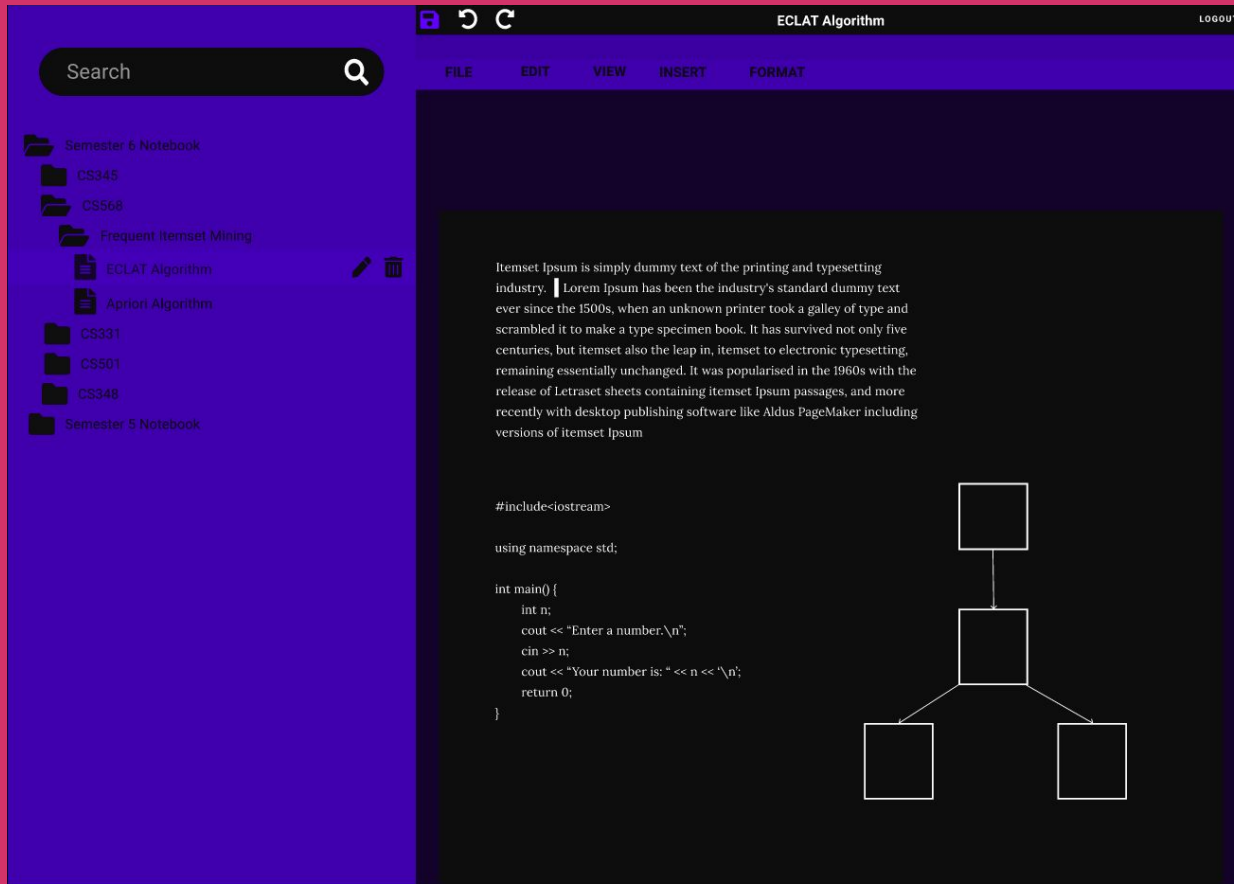
Interface# 4

<Light Theme, Block Cursor>



Interface# 5

<Dark 1 Theme, Block Cursor>



Interface# 6

<Dark 2 Theme, Block Cursor>

Data Collection: Testable Research Question 1

Interface	P#1	P#2	P#3	P#4	P#5	P#6	P#7	P#8	P#9	P#10	P#11	P#12
Lora,11	3	4	3	3	4	5	3	4	4	5	4	6
Lora,14	10	8	10	9	8	8	10	8	10	9	9	8
Lora,17	6	7	8	8	6	7	6	5	6	7	5	6
Nota,11	1	4	3	2	3	4	3	4	3	2	3	3
Nota,14	6	6	5	7	6	5	5	6	6	7	8	6
Nota,17	5	3	5	4	4	3	5	4	2	4	4	5

The following ratings (on a scale of 1-10) were collected from the experiment.



Data Collection: Testable Research Question 2

Interface	P#1	P#2	P#3	P#4	P#5	P#6	P#7	P#8	P#9	P#10	P#11	P#12
Light, Line	10	9	10	8	8	9	10	8	8	9	8	10
Dark1, Line	2	3	3	4	4	3	3	4	4	3	4	6
Dark2, Line	6	7	8	6	7	5	6	7	6	7	5	7
Light, Block	6	5	7	6	8	5	6	6	7	6	7	6
Dark1, Block	2	2	4	3	4	5	4	3	4	3	3	2
Dark2, Block	5	3	5	3	4	6	4	3	4	5	5	4


The following ratings (on a scale of 1-10) were collected from the experiment.



Empirical Data Analysis

It is an important part of the study. In these experiments, we have 12 interfaces in total and 12 participants but we want our conclusions to hold true for any user and any interface. Hence one needs to find out the statistical significance of the obtained data. For this, a parametric test named Repeated Measure Anova has been done. This RMANova test was conducted using the *statsmodel* library in Python.

Repeated measure anova: It fits when the experiment has within-subject design with two or more factors having three or more levels. Since in our experiment both font size in RQ1 and background theme in RQ2 have three levels, it can be applied to our data. Further the three assumptions for a parametric test are verified:-

1. Interval scale having values 1-10 is being used for measurement of the dependent variable i.e Rating.
 2. Variance in the groups of data is approximately equal which is around 0.90 (specifically the range 0.86-0.96) as shown in the analysis.
 3. Data is assumed to come from a normally distributed population (this is an assumption we make which seems reasonable).
- 

Our Results

The results of the repeated measure anova test are shown on the next page. The following is a brief description of what the columns of the table stand for

- F value: F-statistic value for significance
- num DF: Number of degrees of freedom
- den DF: Denominator degrees of freedom
- $Pr > F$: P-value for significance. (smaller means more significant)



Empirical Data Analysis

Statistical Analysis:

Testcase	Mean	Variance
light, line	8.916666667	0.9003366374
dark1, line	3.583333333	0.9962049199
dark2, line	6.416666667	0.9003366374
light, block	6.25	0.8660254038
dark1, block	3.25	0.9653072992
dark2, block	4.25	0.9653072992

Testcase	Mean	Variance
11, Lora	4	0.9534625892
14, Lora	8.916666667	0.9003366374
17, Lora	6.416666667	0.9962049199
11, Nota	2.916666667	0.9003366374
14, Nota	6.083333333	0.9003366374
17, Nota	4	0.9534625892



Our Results

```
In [14]: AnovaRMTable('anova5.csv', 'font_size', 'font_style')
```

```
=====
                        Anova
=====
      F Value  Num DF  Den DF  Pr > F
-----
font_size      244.0586  2.0000 20.0000 0.0000
font_style      32.9513  1.0000 10.0000 0.0002
font_size:font_style  7.8560  2.0000 20.0000 0.0030
=====
```

```
In [15]: AnovaRMTable('anova6.csv', var1='background_theme', var2='cursor_style')
```

```
=====
                        Anova
=====
      F Value  Num DF  Den DF  Pr > F
-----
background theme      71.8504  2.0000 20.0000 0.0000
cursor_style          107.7149  1.0000 10.0000 0.0000
background_theme:cursor_style  5.5802  2.0000 20.0000 0.0119
=====
```

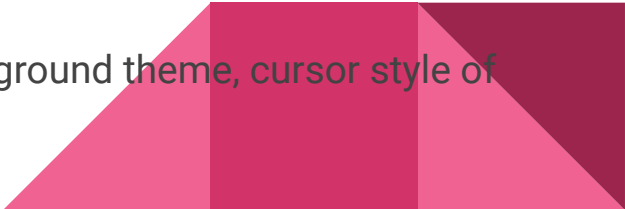
Conclusions

The results show the following:

1. $F(2) = 7.856$, $p < 0.05$, statistically significant.
2. $F(2) = 5.580$, $p < 0.05$, statistically significant.

The $Pr > F$ values that act as p-values are all less than 0.05, in the table show that the experimental results are significant for both experiments and the probability of getting the results by chance is less than 5%.

Hence, we reject both null hypotheses given below:

1. Aesthetic score (in a scale of 1-10) does not depend on the font size, font type of the note interface.
 2. Aesthetic score (in a scale of 1-10) does not depend on the background theme, cursor style of the note interface.
- 

Conclusions

The means obtained from the experimental data clearly indicate that the ideal setting for the default note interface is 14 pts text size, Lora (Serif) font face with light background and blinking line cursor. The statistical significance tests support these finding.





Thank you!