Project Title: Student Budget Planner – Analyzing College Student Spending Patterns

---

Objective:

To analyze the financial behavior of college students by exploring their income, expenses, and savings trends using data-driven tools. The goal is to uncover key insights about budgeting challenges and spending distribution to improve financial planning awareness among students.

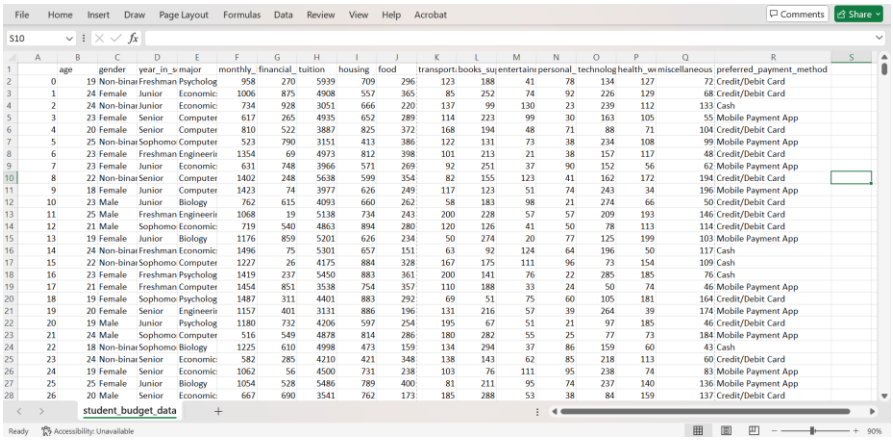
Tools Used:

- **Excel** – Data cleaning, formulas, dashboard creation
- **SQL (MySQL Workbench)** – Table creation and query execution
- **Tableau Public** – Interactive dashboard visualizations
- **Python (Jupyter Notebook)** – EDA, plots, and correlation analysis
- **Pandas, Seaborn, Matplotlib** – For data handling & visualization

---

Dataset Overview:

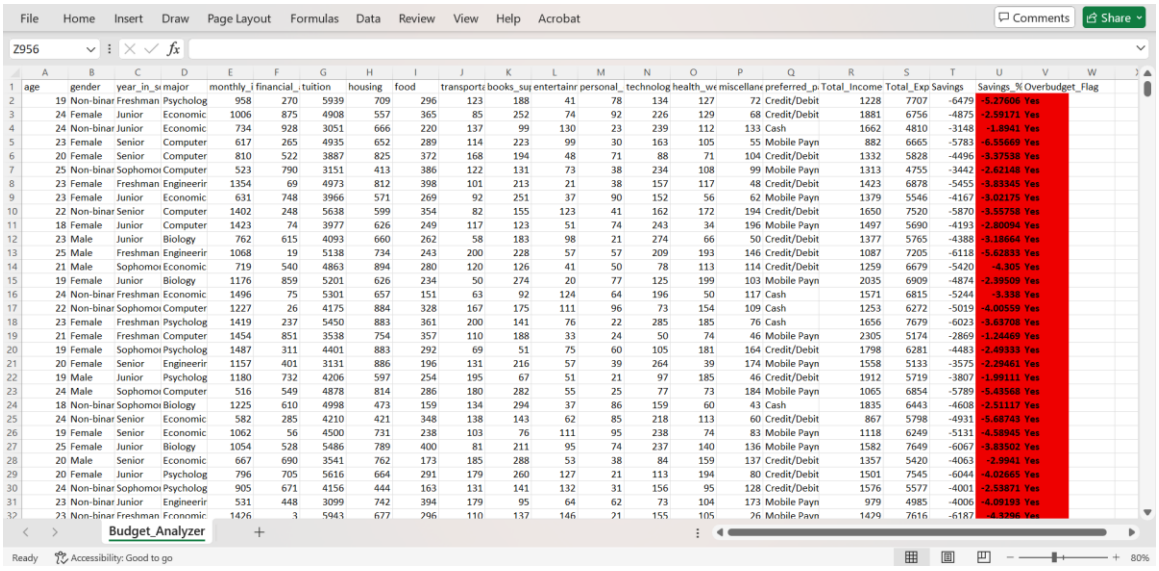
- **File:** student_budget_data.csv
- **Records:** 1000 students
- **Fields:**
    - Demographics: major, monthly_income, financial_aid
    - Expenses: tuition, housing, food, transportation, etc.
    - Derived columns: total_expense, savings, savings_percent, overbudget_flag



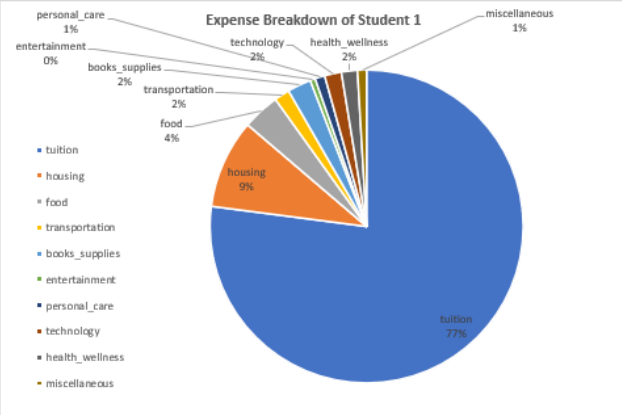
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		age	gender	year_in_s	major	monthly_	financial_	tuition	housing	food	transport	books_su	entertain	personal_	technolog	health_w	miscellaneous	preferred_payment_method	
2		0	19	Non-binar	Freshman	Psycholog	958	270	5939	709	296	123	188	41	78	134	127	72	Credit/Debit Card	
3		1	24	Female	Junior	Economic	1006	875	4908	557	365	85	252	74	92	226	129	68	Credit/Debit Card	
4		2	24	Non-binar	Junior	Economic	734	928	3051	666	220	137	99	130	23	239	112	133	Cash	
5		3	23	Female	Senior	Computer	617	265	4935	652	289	114	223	99	30	163	105	55	Mobile Payment App	
6		4	20	Female	Senior	Computer	810	522	3887	825	372	168	194	48	71	88	71	104	Credit/Debit Card	
7		5	25	Non-binar	Sophomo	Computer	523	790	3151	413	386	122	131	73	38	234	108	99	Mobile Payment App	
8		6	23	Female	Freshman	Engineeri	1354	69	4973	812	398	101	213	21	38	157	117	48	Credit/Debit Card	
9		7	23	Female	Junior	Economic	631	748	3966	571	269	92	251	37	90	152	56	62	Mobile Payment App	
10		8	22	Non-binar	Senior	Computer	1402	248	5638	599	354	82	155	123	41	162	172	194	Credit/Debit Card	
11		9	18	Female	Junior	Computer	1423	74	3977	626	249	117	123	51	74	243	34	196	Mobile Payment App	
12		10	23	Male	Junior	Biology	762	615	4093	660	262	58	183	98	21	274	66	50	Credit/Debit Card	
13		11	25	Male	Freshman	Engineeri	1068	19	5138	734	243	200	228	57	57	209	193	146	Credit/Debit Card	
14		12	21	Male	Sophomo	Economic	719	540	4863	894	280	120	126	41	50	78	113	114	Credit/Debit Card	
15		13	19	Female	Junior	Biology	1176	859	5201	626	234	50	274	20	77	125	199	103	Mobile Payment App	
16		14	24	Non-binar	Freshman	Economic	1496	75	5301	657	151	63	92	124	64	196	50	117	Cash	
17		15	22	Non-binar	Sophomo	Computer	1227	26	4175	884	328	167	175	111	96	73	154	109	Cash	
18		16	23	Female	Freshman	Psycholog	1419	237	5450	883	361	200	141	76	22	285	185	76	Cash	
19		17	21	Female	Freshman	Computer	1454	851	3538	754	357	110	188	33	24	50	74	46	Mobile Payment App	
20		18	19	Female	Sophomo	Psycholog	1487	311	4401	883	292	69	51	75	60	105	181	164	Credit/Debit Card	
21		19	20	Female	Senior	Engineeri	1157	401	3131	886	196	131	216	57	39	264	39	174	Mobile Payment App	
22		20	19	Male	Junior	Psycholog	1180	732	4206	597	254	195	67	51	21	97	185	46	Credit/Debit Card	
23		21	24	Male	Sophomo	Computer	516	549	4878	814	286	180	282	55	25	77	73	184	Mobile Payment App	
24		22	18	Non-binar	Sophomo	Biology	1225	610	4998	473	159	134	294	37	86	159	60	43	Cash	
25		23	24	Non-binar	Senior	Economic	582	285	4210	421	348	138	143	62	85	218	113	60	Credit/Debit Card	
26		24	19	Female	Senior	Economic	1062	56	4500	731	238	103	76	111	95	238	74	83	Mobile Payment App	
27		25	25	Female	Junior	Biology	1054	528	5486	789	400	81	211	95	74	237	140	136	Mobile Payment App	
28		26	20	Male	Senior	Economic	667	690	3541	762	173	185	288	53	38	84	159	137	Credit/Debit Card	

Excel Work Summary:

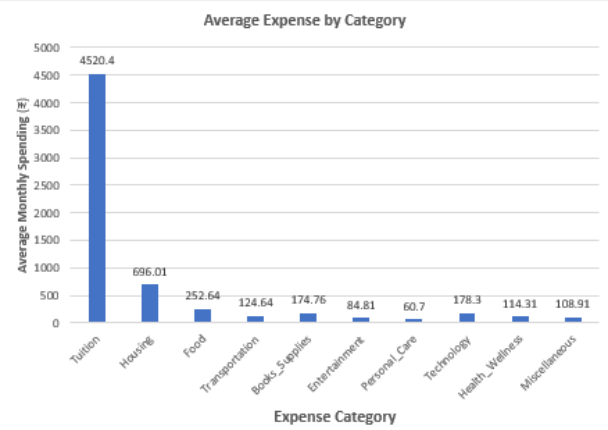
- Cleaned dataset in Budget_Analyzer.xlsx
- Created calculated columns:
    - Total_Income = monthly_income + financial_aid
    - Total_Expense = SUM(all expenses)
    - Savings = Total_Income - Total_Expense
    - Savings_% = Savings / Total_Income
    - Overbudget Flag = Yes/No
- Create Conditional Formating:
    - Savings_%
    - Overbudget_Flag



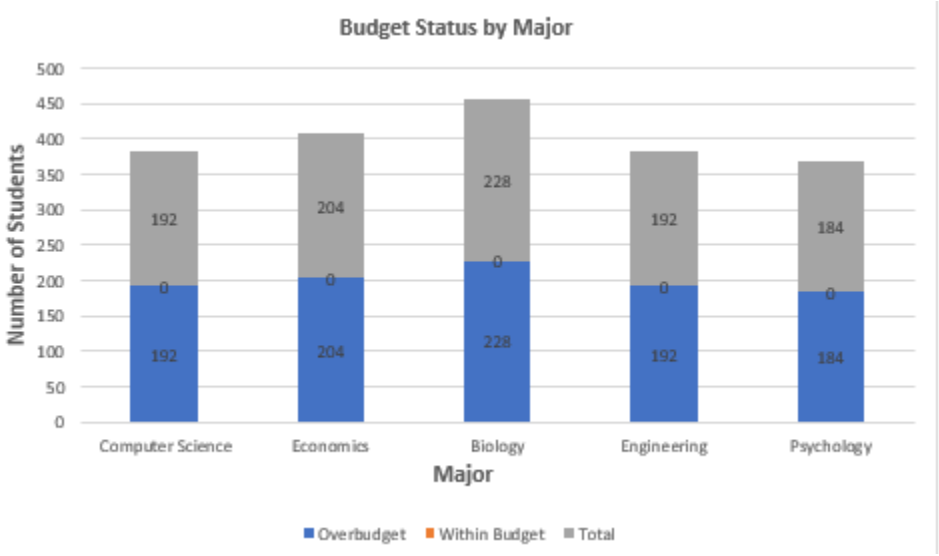
- Built chart in Budget_Calculator.xlsx with:
    - Expence Breakdown of Student 1



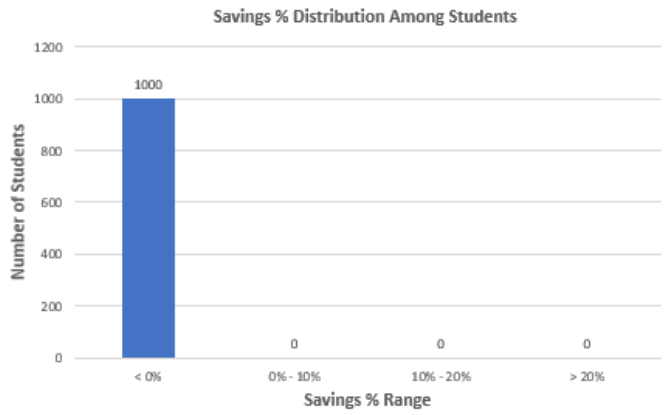
    - Average Expense by Category



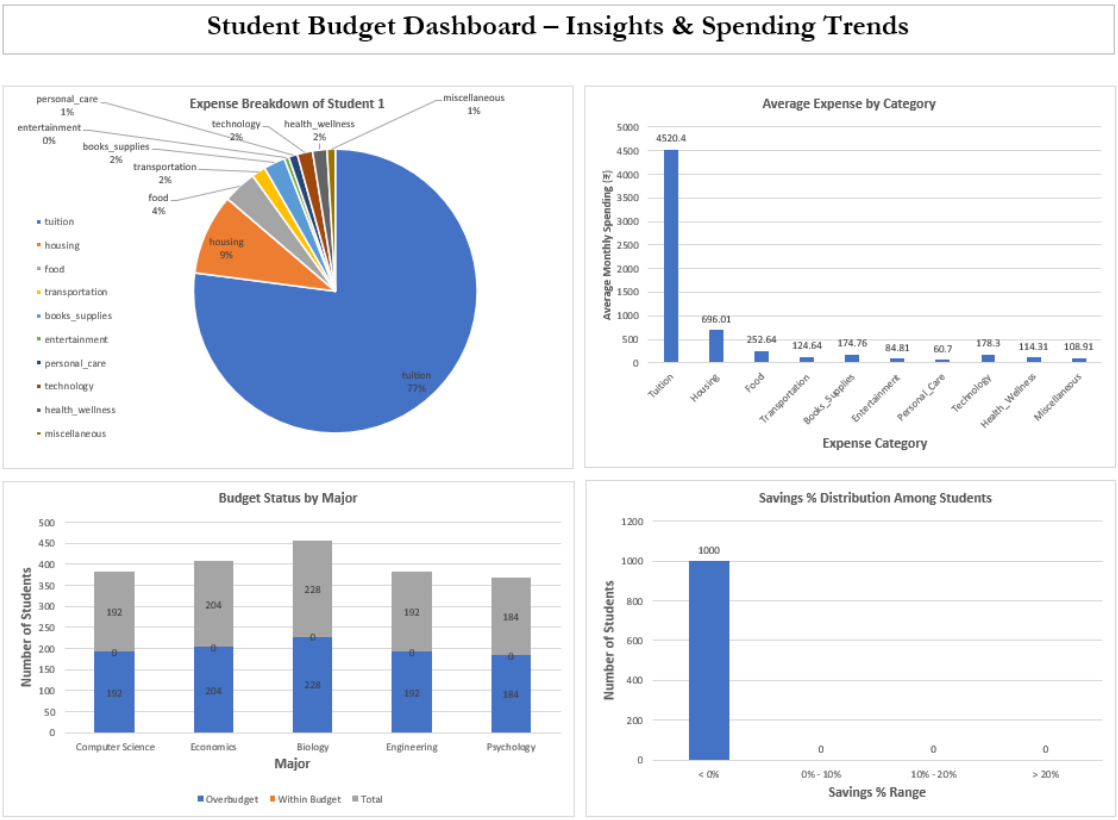
    - Budges Status by Major



- Savings % Distribution Among Students



- **Final Dashboard**

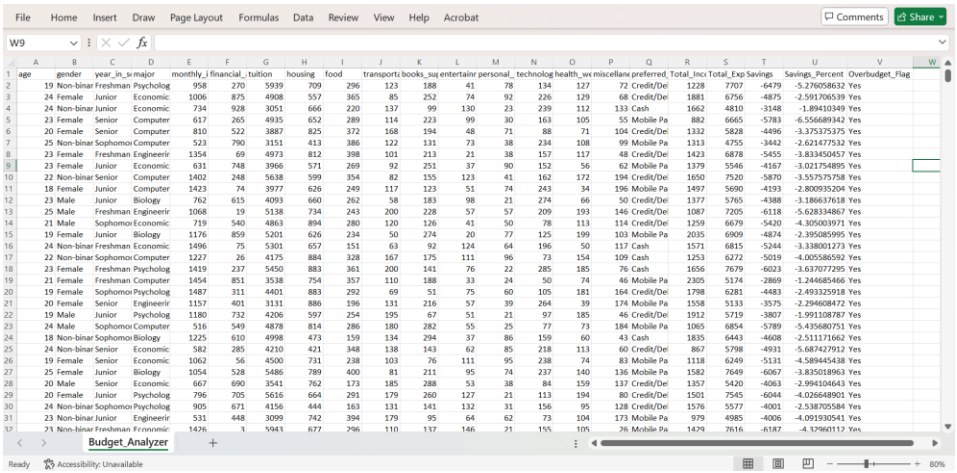


SQL Work Summary:

CREATE DATABASE student_budget;

USE student_budget;

- Imported cleaned .csv using MySQL Workbench



- Created schema in schema.sql
- Planned to run queries (to be added in queries.sql) like:

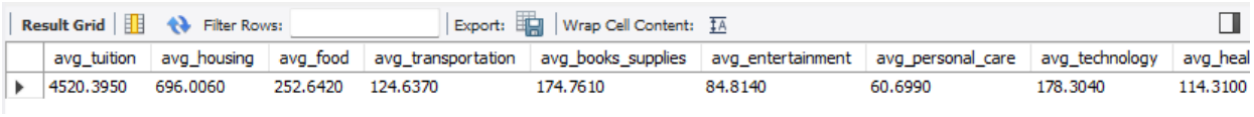
```
-- Top 10 Students with Highest Savings
SELECT age, gender, major, total_income, total_expense, savings
FROM student_budget.budget_analyzer
ORDER BY savings DESC
LIMIT 10;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

age	gender	major	total_income	total_expense	savings
21	Female	Biology	2308	4583	-2275
23	Female	Biology	2180	4542	-2362
19	Male	Economics	2251	4712	-2461
25	Female	Economics	2042	4612	-2570
23	Non-binary	Biology	1880	4453	-2573
25	Non-binary	Biology	2309	4886	-2577
24	Non-binary	Biology	2427	5028	-2601

```
-- Average Expense by Category
SELECT
    AVG(tuition) AS avg_tuition,
    AVG(housing) AS avg_housing,
    AVG(food) AS avg_food,
    AVG(transportation) AS avg_transportation,
```

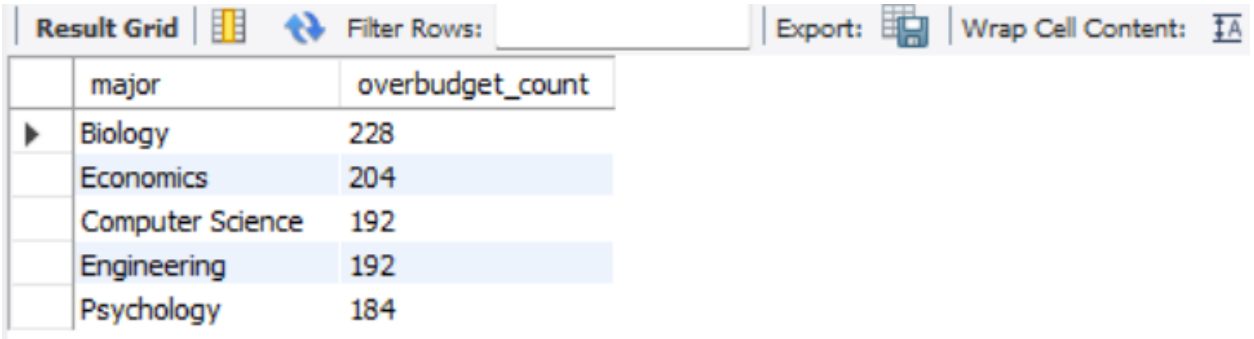
```
    AVG(books_supplies) AS avg_books_supplies,
    AVG(entertainment) AS avg_entertainment,
    AVG(personal_care) AS avg_personal_care,
    AVG(technology) AS avg_technology,
    AVG(health_wellness) AS avg_health_wellness,
    AVG(miscellaneous) AS avg_miscellaneous
FROM student_budget.budget_analyzer;
```



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

avg_tuition	avg_housing	avg_food	avg_transportation	avg_books_supplies	avg_entertainment	avg_personal_care	avg_technology	avg_heal
4520.3950	696.0060	252.6420	124.6370	174.7610	84.8140	60.6990	178.3040	114.3100

```
-- Total Overbudget Students by Major
SELECT major, COUNT(*) AS overbudget_count
FROM student_budget.budget_analyzer
WHERE overbudget_flag = 'Yes'
GROUP BY major
ORDER BY overbudget_count DESC;
```

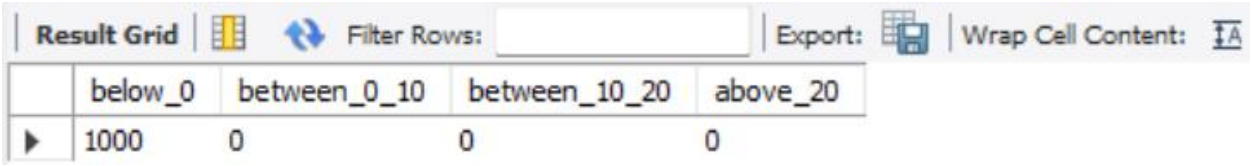


Result Grid | Filter Rows: | Export: | Wrap Cell Content:

major	overbudget_count
Biology	228
Economics	204
Computer Science	192
Engineering	192
Psychology	184

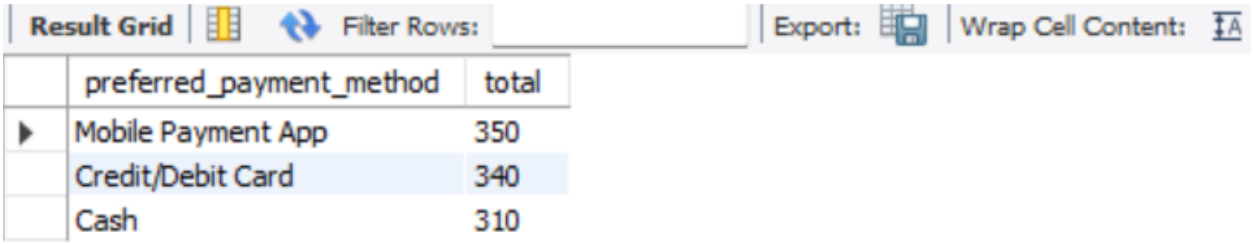
```
-- Savings % Distribution
SELECT
    SUM(CASE WHEN savings_percent < 0 THEN 1 ELSE 0 END) AS below_0,
    SUM(CASE WHEN savings_percent BETWEEN 0 AND 0.1 THEN 1 ELSE 0 END) AS
between_0_10,
    SUM(CASE WHEN savings_percent > 0.1 AND savings_percent <= 0.2 THEN 1 ELSE 0
END) AS between_10_20,
```

```
    SUM(CASE WHEN savings_percent > 0.2 THEN 1 ELSE 0 END) AS above_20
FROM student_budget.budget_analyzer;
```



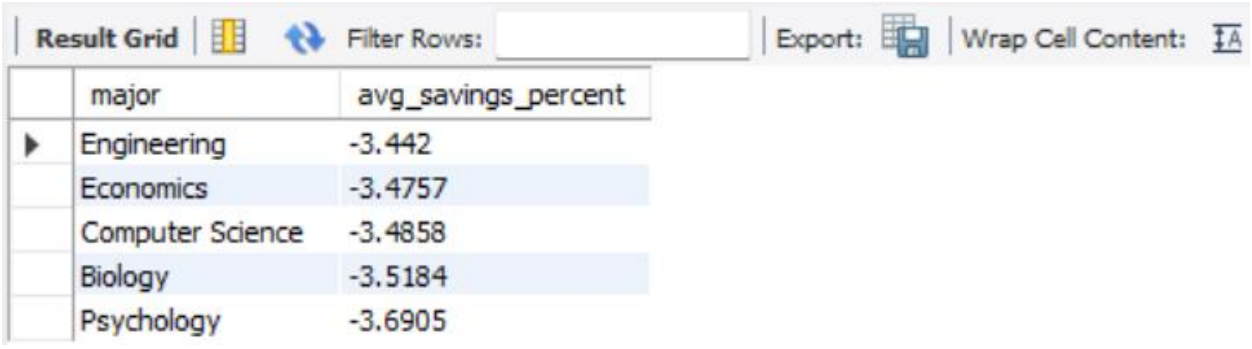
	below_0	between_0_10	between_10_20	above_20
▶	1000	0	0	0

```
-- Count of Students by Preferred Payment Method
SELECT preferred_payment_method, COUNT(*) AS total
FROM student_budget.budget_analyzer
GROUP BY preferred_payment_method
ORDER BY total DESC;
```



	preferred_payment_method	total
▶	Mobile Payment App	350
	Credit/Debit Card	340
	Cash	310

```
-- Major-wise Average Savings %
SELECT major, ROUND(AVG(savings_percent), 4) AS avg_savings_percent
FROM student_budget.budget_analyzer
GROUP BY major
ORDER BY avg_savings_percent DESC;
```

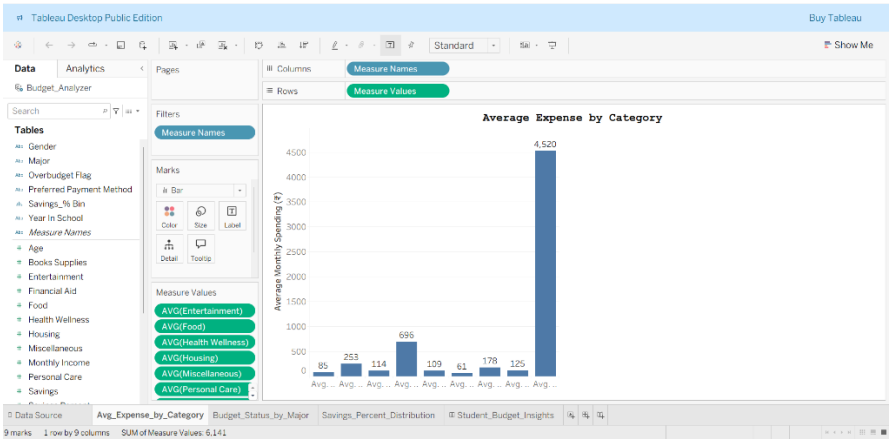


	major	avg_savings_percent
▶	Engineering	-3.442
	Economics	-3.4757
	Computer Science	-3.4858
	Biology	-3.5184
	Psychology	-3.6905

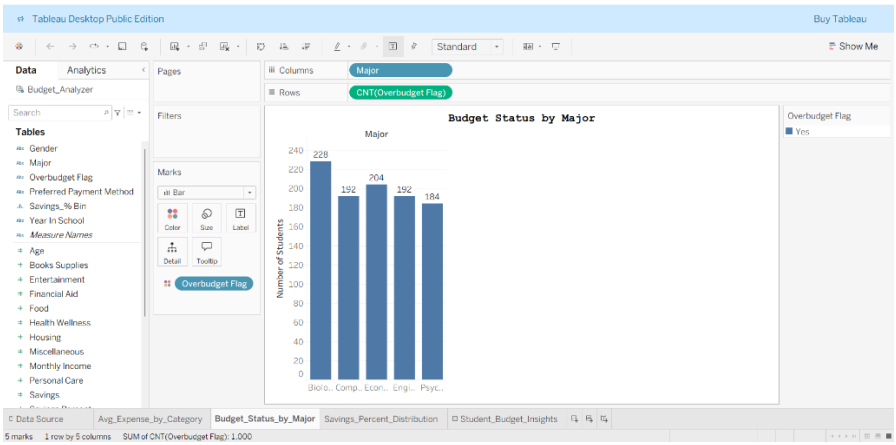
Tableau Work Summary:

Dashboard created in Tableau Public with:

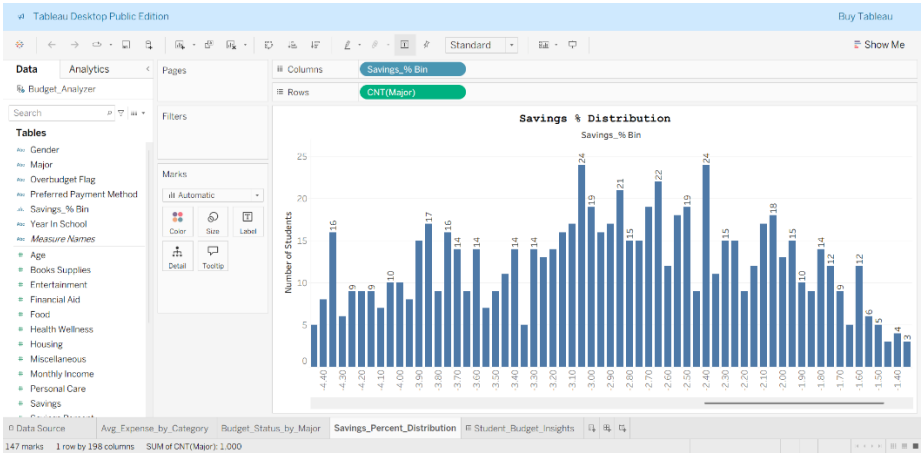
- **Chart 1:** Average Expense by Category



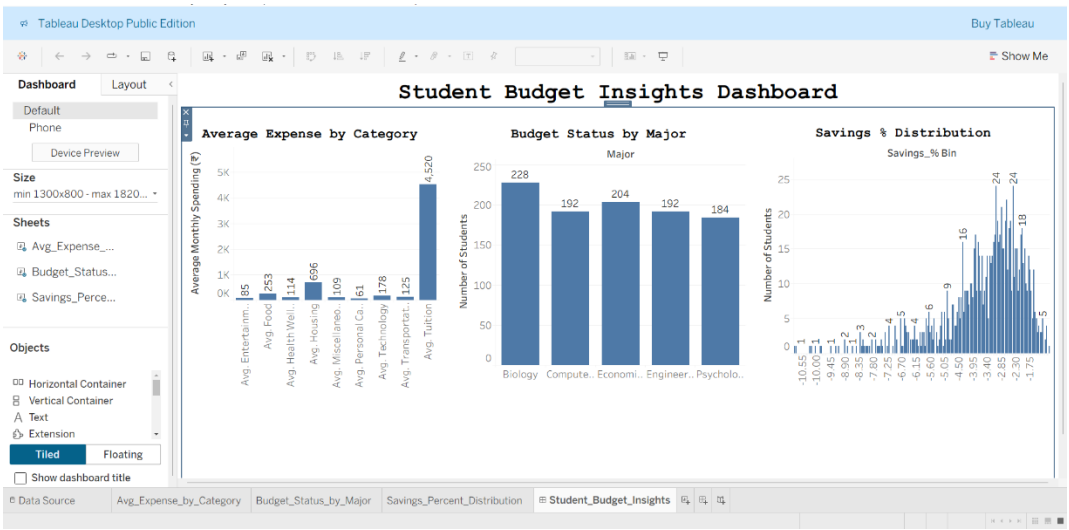
- **Chart 2:** Budget Status by Major



- **Chart 3:** Savings % Distribution



- **Final Dashboard**



- Published as: dashboard.twbx

Python Work Summary:

Jupyter notebook: analysis.ipynb

Import Libraries

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

Load Data

df = pd.read_csv(r'E:/Student_Budget_Planner_Project/excel/Budget_Analyzer.csv')

df.head()

	age	gender	year_in_school	major	monthly_income	financial_aid	tuition	housing	food	transportation	...	personal_care	technology	health_wellness	misce
0	19	Non-binary	Freshman	Psychology	958	270	5939	709	296	123	...	78	134	127	
1	24	Female	Junior	Economics	1006	875	4908	557	365	85	...	92	226	129	
2	24	Non-binary	Junior	Economics	734	928	3051	666	220	137	...	23	239	112	
3	23	Female	Senior	Computer Science	617	265	4935	652	289	114	...	30	163	105	
4	20	Female	Senior	Computer Science	810	522	3887	825	372	168	...	71	88	71	

5 rows × 22 columns

Basic Overview

df.info()

df.describe()

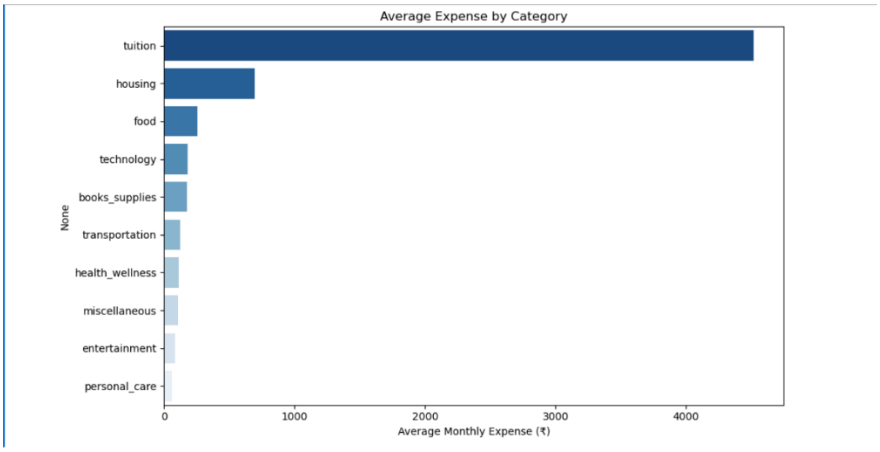
df.isnull().sum()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 22 columns):
 #   Column                     Non-Null Count  Dtype  
---  ------                     --------------  -----  
 0   age                        1000 non-null   int64  
 1   gender                     1000 non-null   object 
 2   year_in_school             1000 non-null   object 
 3   major                      1000 non-null   object 
 4   monthly_income             1000 non-null   int64  
 5   financial_aid              1000 non-null   int64  
 6   tuition                    1000 non-null   int64  
 7   housing                    1000 non-null   int64  
 8   food                       1000 non-null   int64  
 9   transportation             1000 non-null   int64  
 10  books_supplies             1000 non-null   int64  
 11  entertainment              1000 non-null   int64  
 12  personal_care              1000 non-null   int64  
 13  technology                 1000 non-null   int64  
 14  health_wellness            1000 non-null   int64  
 15  miscellaneous              1000 non-null   int64  
 16  preferred_payment_method   1000 non-null   object 
 17  Total_Income               1000 non-null   int64  
 18  Total_Expense              1000 non-null   int64  
 19  Savings                    1000 non-null   int64  
 20  Savings_Percent            1000 non-null   float64
 21  Overbudget_Flag            1000 non-null   object 
dtypes: float64(1), int64(16), object(5)
memory usage: 172.0+ KB
```

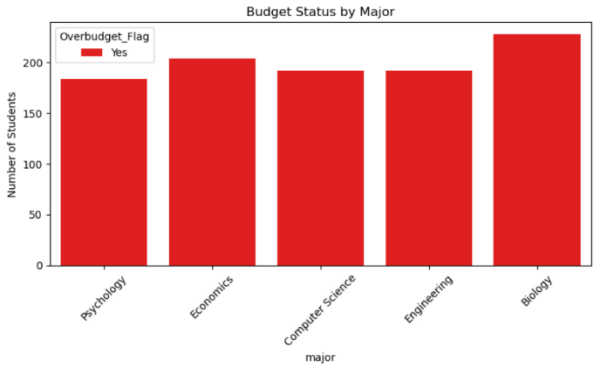
```
age                         0
gender                      0
year_in_school              0
major                       0
monthly_income              0
financial_aid               0
tuition                     0
housing                     0
food                        0
transportation              0
books_supplies              0
entertainment               0
personal_care               0
technology                  0
health_wellness             0
miscellaneous               0
preferred_payment_method    0
Total_Income                0
Total_Expense               0
Savings                     0
Savings_Percent             0
Overbudget_Flag             0
dtype: int64
```

Visualizations

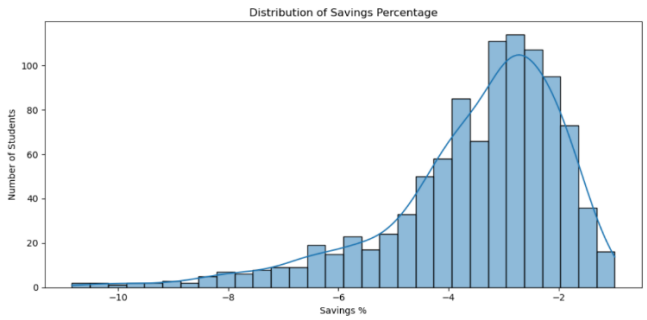
```
# Average Expense by Category
expense_cols = ['tuition', 'housing', 'food', 'transportation', 'books_supplies',
                'entertainment', 'personal_care', 'technology', 'health_wellness', 'miscellaneous']

avg_expense = df[expense_cols].mean().sort_values(ascending=False)

plt.figure(figsize=(10,6))
sns.barplot(x=avg_expense.values, y=avg_expense.index, palette='Blues_r')
plt.xlabel("Average Monthly Expense (₹)")
plt.title("Average Expense by Category")
plt.tight_layout()
plt.show()
```



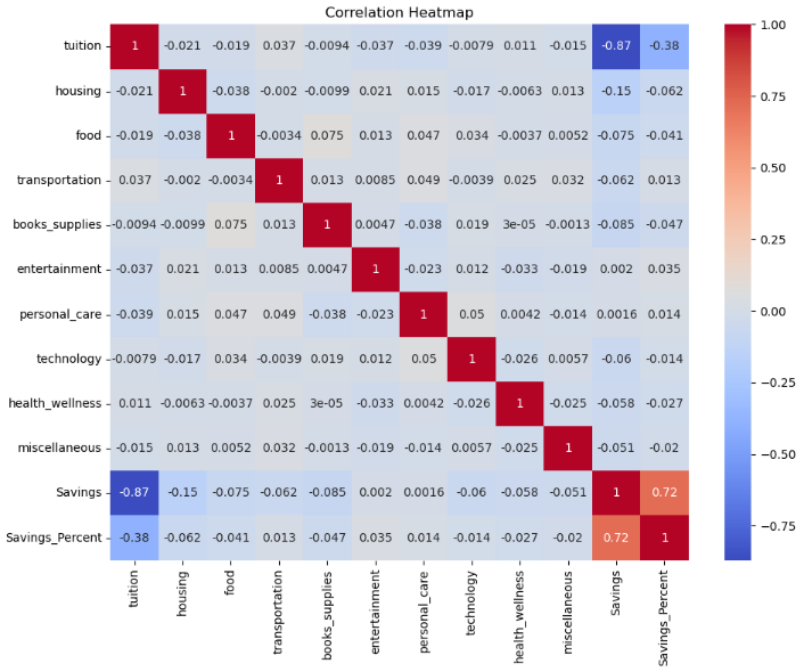
```
# Budget Status by Major
plt.figure(figsize=(8,5))
sns.countplot(data=df, x='major', hue='Overbudget_Flag', palette={'Yes':'red', 'No':'green'})
plt.title("Budget Status by Major")
plt.ylabel("Number of Students")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```




```
# Savings % Distribution
plt.figure(figsize=(10,5))
sns.histplot(df['Savings_Percent'], bins=30, kde=True)
plt.title("Distribution of Savings Percentage")
plt.xlabel("Savings %")
plt.ylabel("Number of Students")
plt.tight_layout()
plt.show()
```



```
# Correlation Heatmap
plt.figure(figsize=(10,8))
sns.heatmap(df[expense_cols + ['Savings', 'Savings_Percent']].corr(), annot=True, cmap='coolwarm')
plt.title("Correlation Heatmap")
plt.tight_layout()
plt.show()
```



Insights & Summary:

- **High tuition & housing costs** are main drivers of overbudgeting
- **More than 90%** students are spending more than their income
- **Biology majors** had the highest overbudget count
- **Financial aid** doesn't fully cover expense burden for most

---

Conclusion:

This project successfully tracks and visualizes key student budgeting patterns. It provides insight into overspending behaviors, allowing institutions or students to take corrective financial actions. It also showcases a complete end-to-end data analysis lifecycle using multiple tools, making it an ideal addition to a data analytics portfolio.