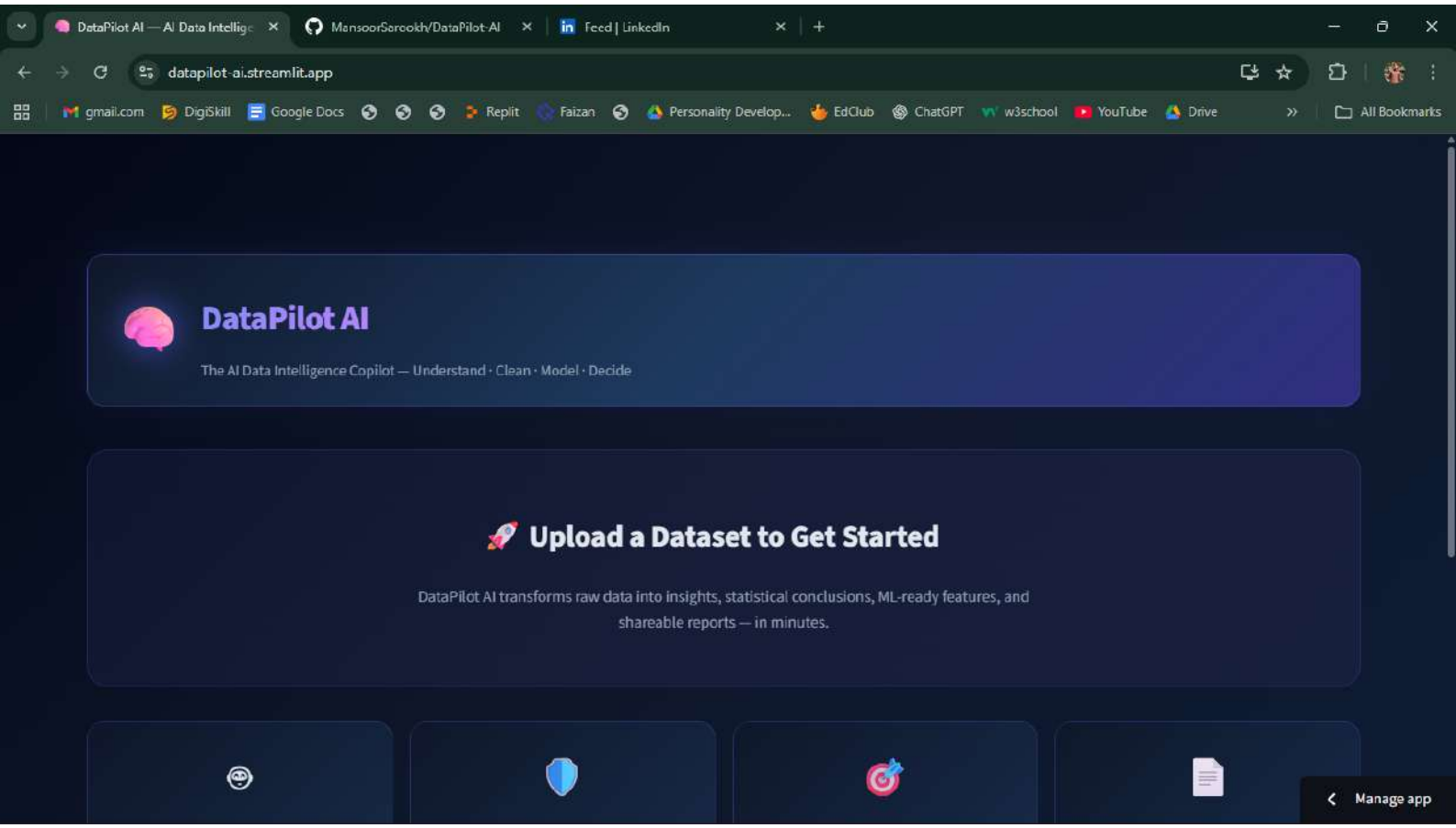
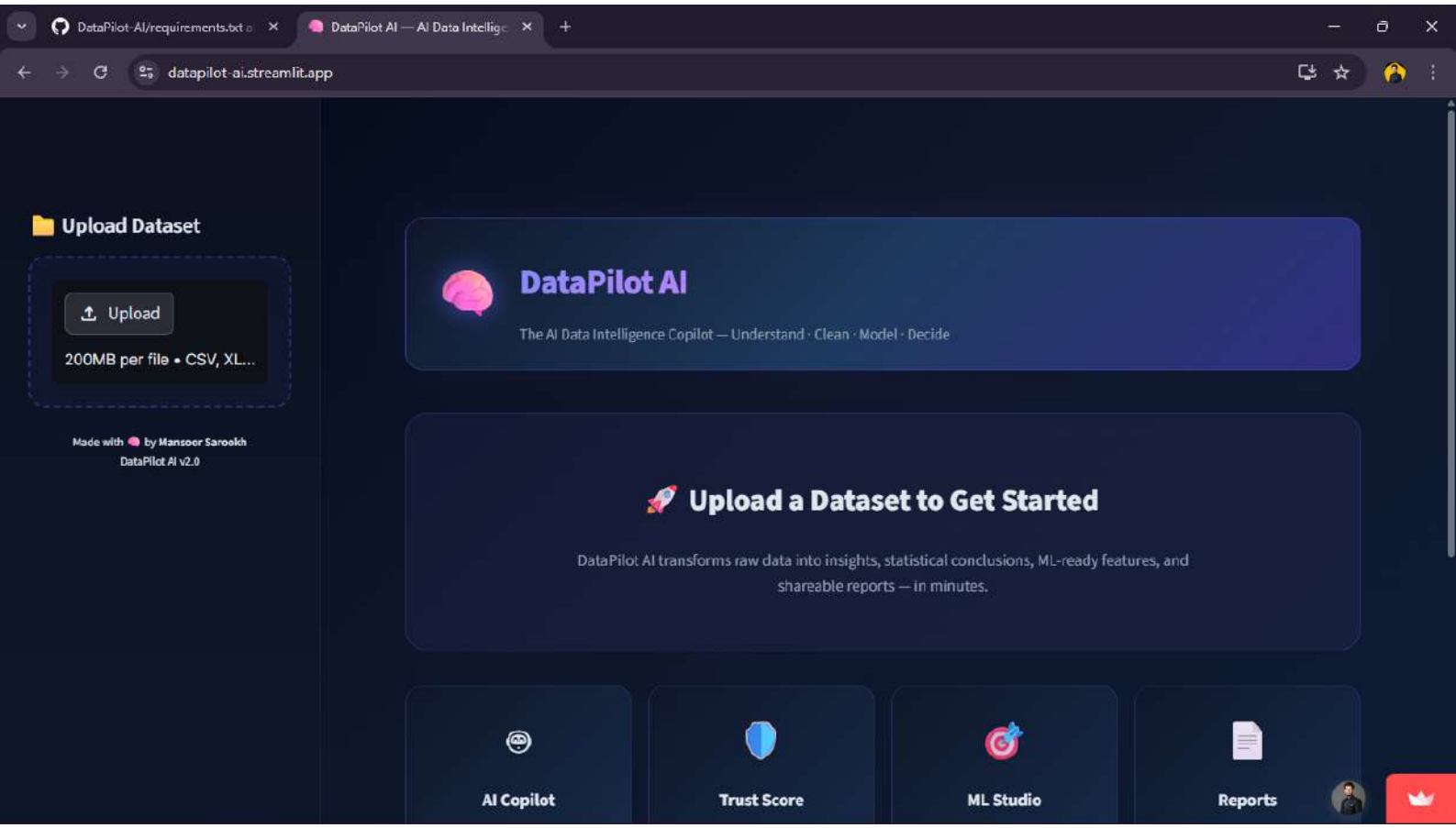
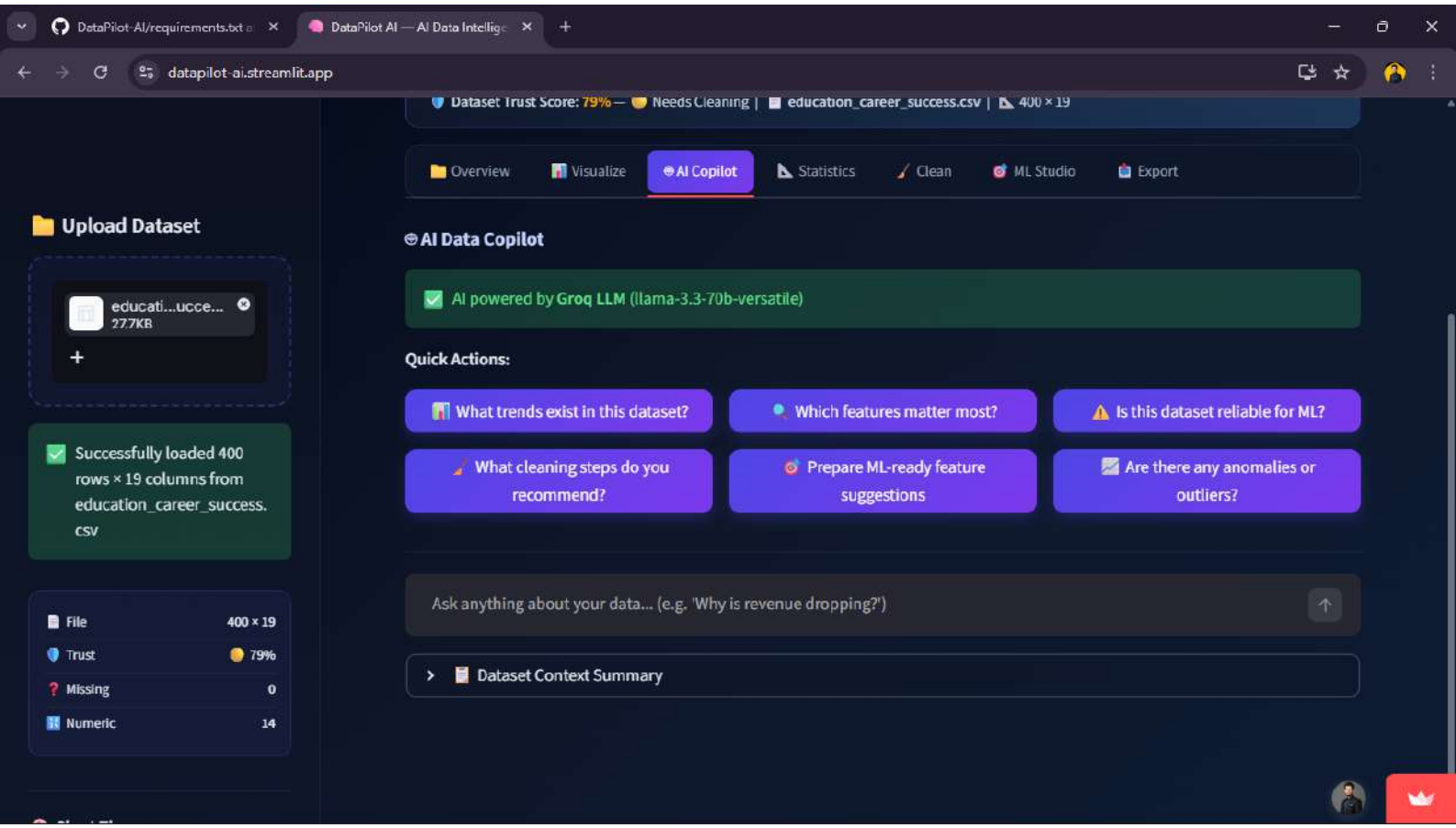
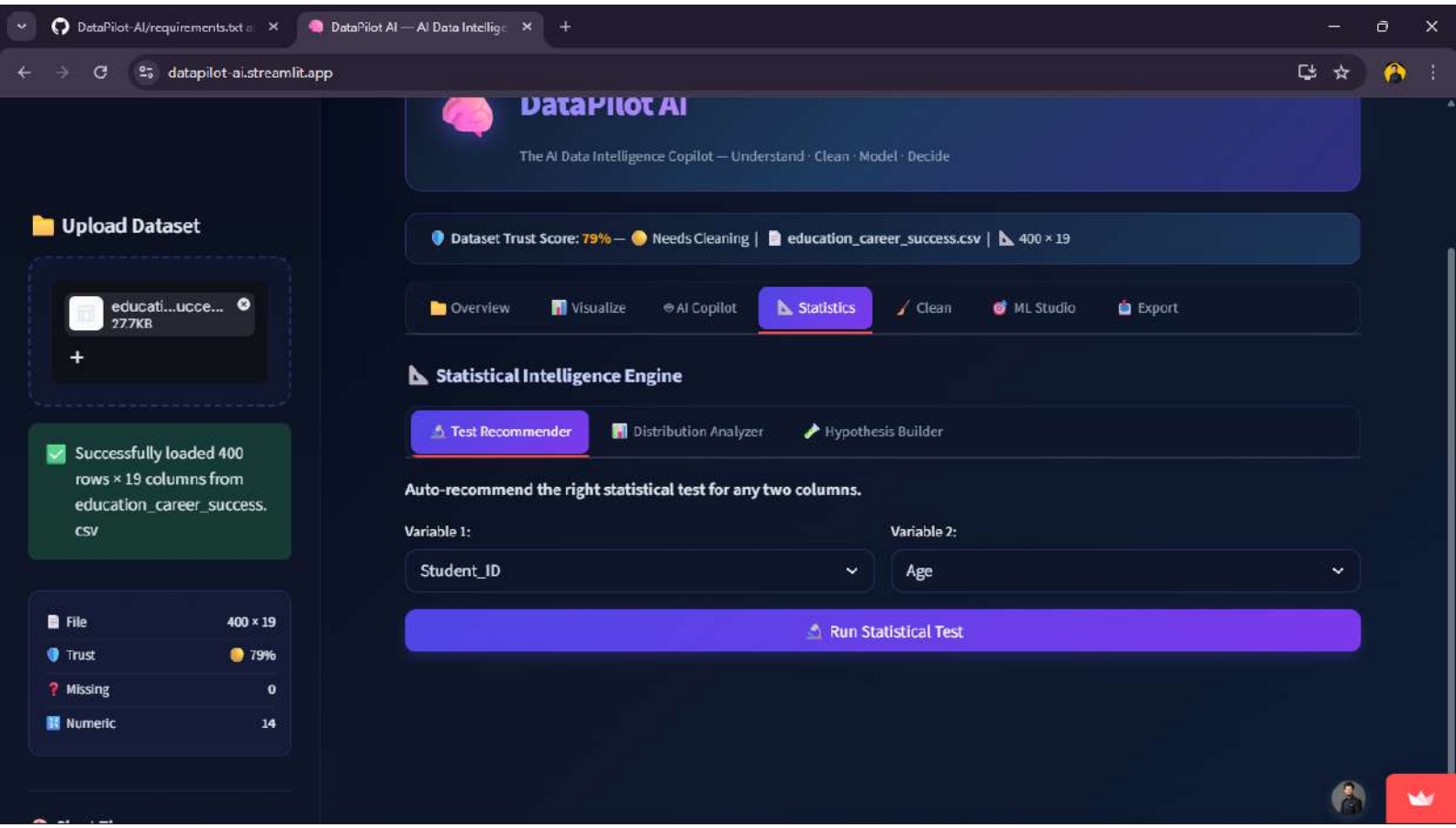
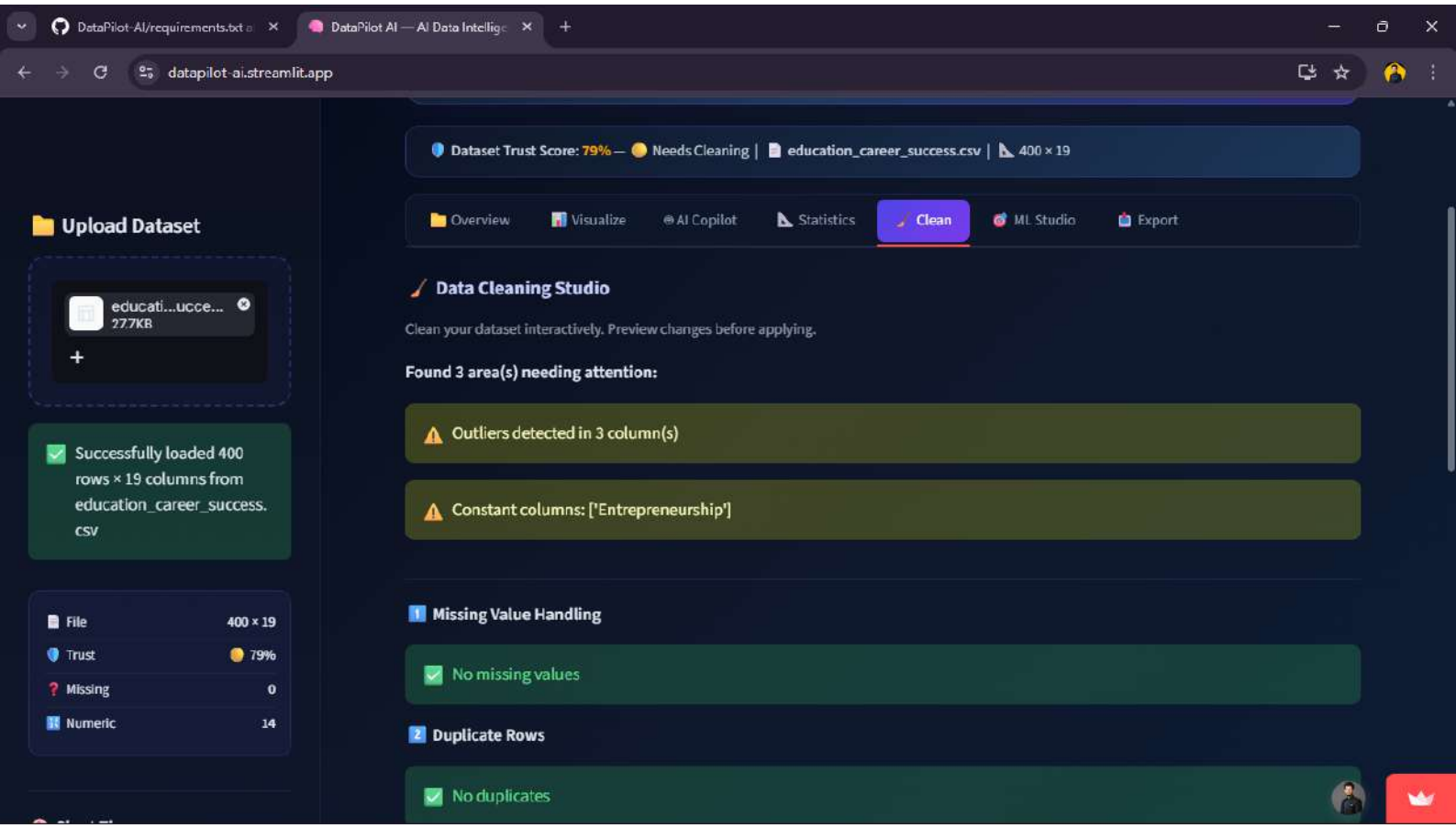
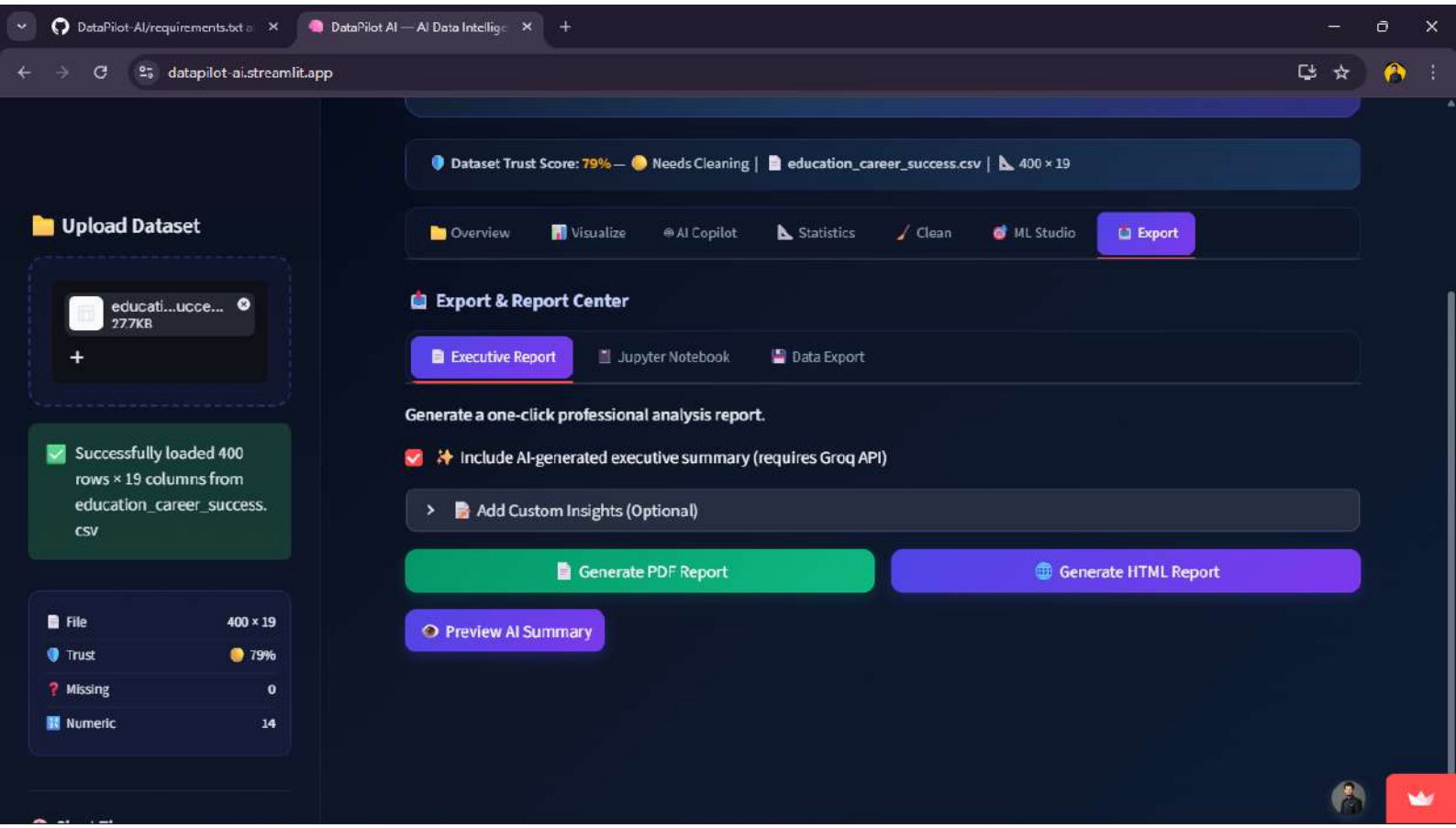

DataPilot AI — AI Data Intellig
MansoorSarookh/DataPilot-AI
Feed | LinkedIn
datapilot-ai.streamlit.app
gmail.com
DigiSkill
Google Docs
Replit
Faizan
Personality Develop...
EdClub
ChatGPT
w3school
YouTube
Drive
All Bookmarks
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
🚀 Upload a Dataset to Get Started
DataPilot AI transforms raw data into insights, statistical conclusions, ML-ready features, and shareable reports — in minutes.
Manage app


DataPilot-AI/requirements.txt
DataPilot AI — AI Data Intellig
datapilot-ai.streamlit.app
Upload Dataset
Upload
200MB per file • CSV, XL...
Made with by Mansoor Sarookh
DataPilot AI v2.0
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Upload a Dataset to Get Started
DataPilot AI transforms raw data into insights, statistical conclusions, ML-ready features, and shareable reports — in minutes.
AI Copilot
Trust Score
ML Studio
Reports

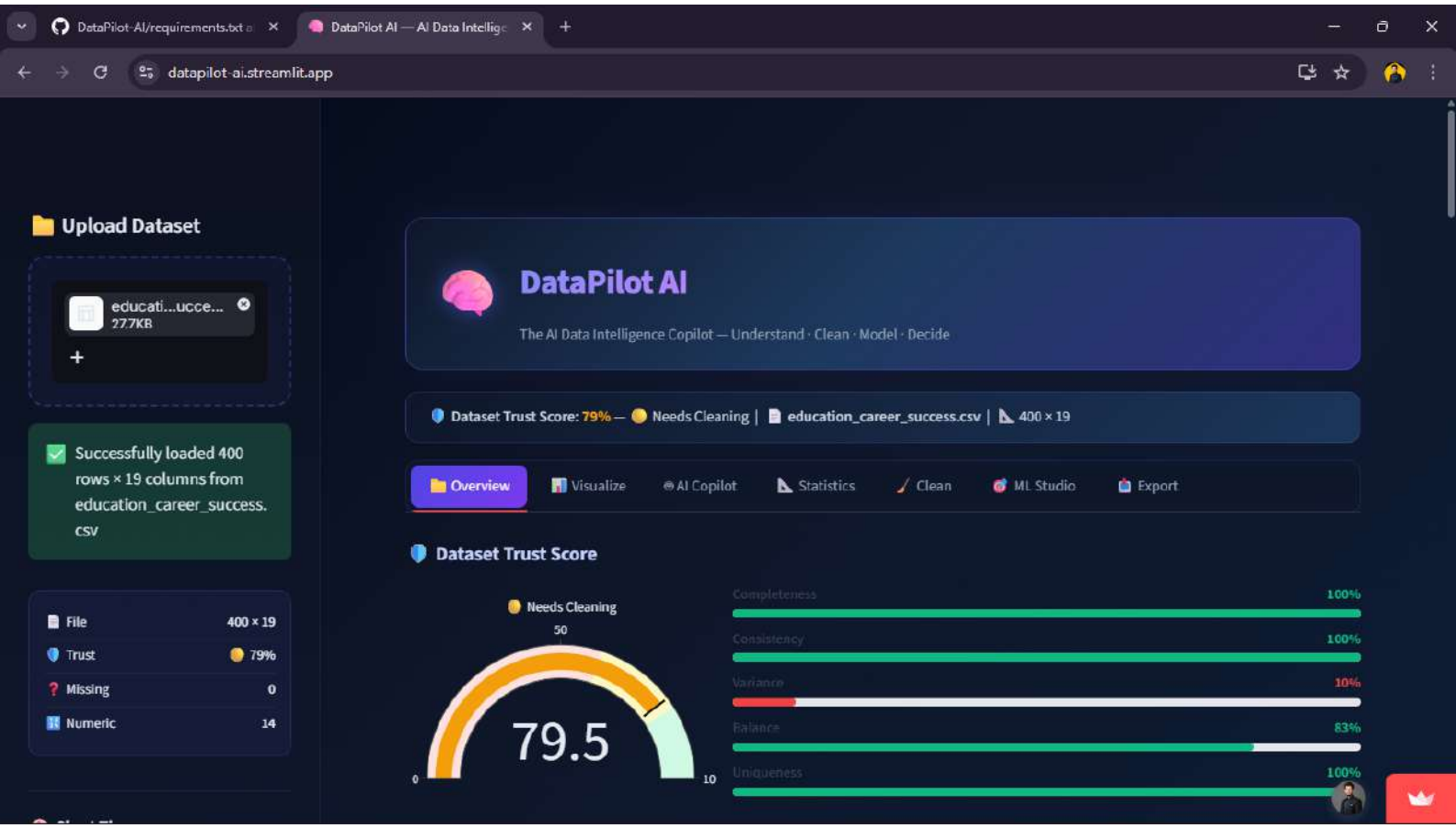

DataPilot-AI/requirements.txt
DataPilot AI — AI Data Intellige
datapilot-ai.streamlit.app
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Upload Dataset
educati...ucce...
27.7KB
Successfully loaded 400 rows × 19 columns from education_career_success.csv
File 400 × 19
Trust 79%
Missing 0
Numeric 14
AI Data Copilot
AI powered by Groq LLM (llama-3.3-70b-versatile)
Quick Actions:
What trends exist in this dataset?
Which features matter most?
Is this dataset reliable for ML?
What cleaning steps do you recommend?
Prepare ML-ready feature suggestions
Are there any anomalies or outliers?
Ask anything about your data... (e.g. 'Why is revenue dropping?')
Dataset Context Summary

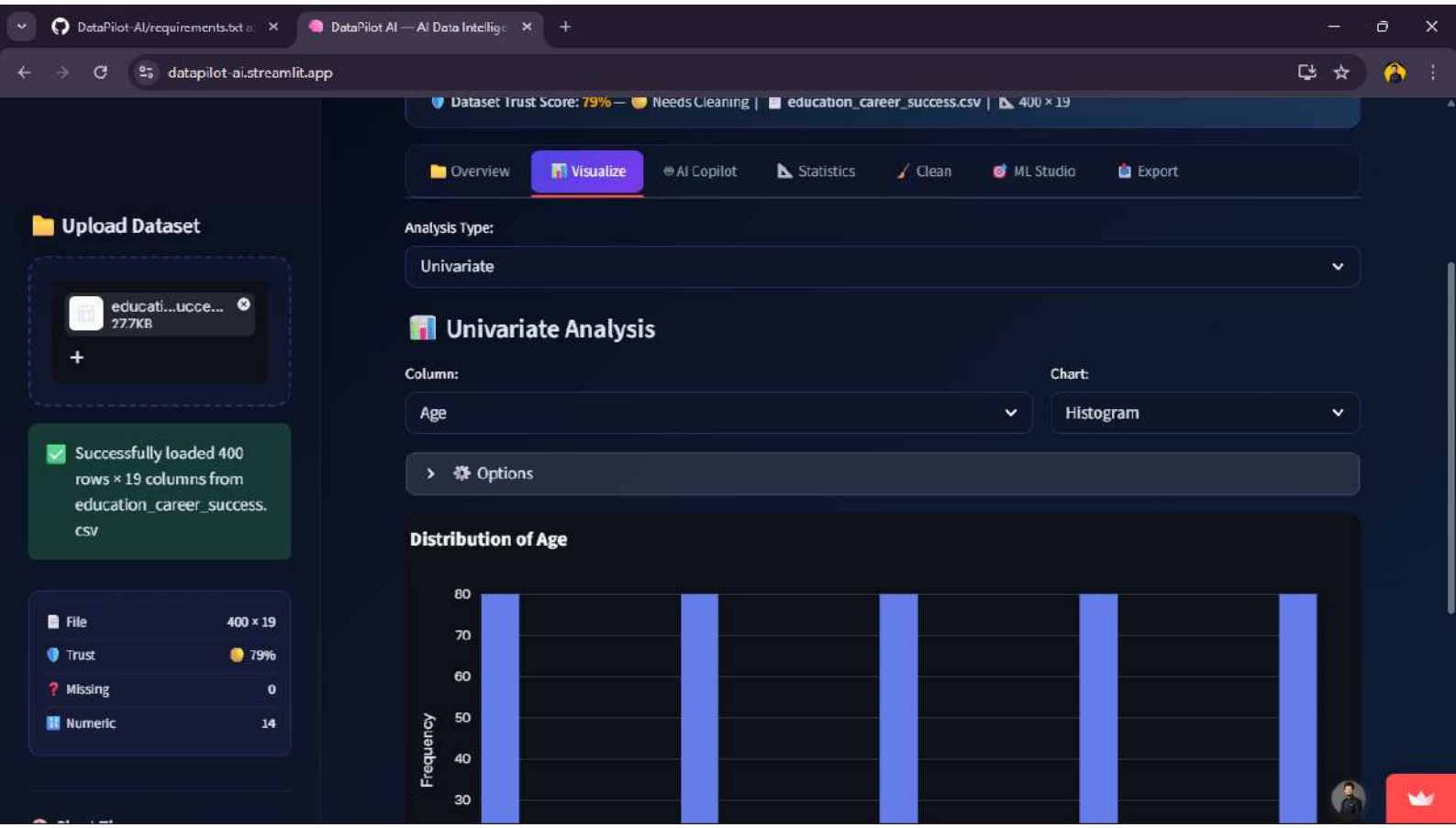

DataPilot-AI/requirements.txt
DataPilot AI — AI Data Intellig
datapilot-ai.streamlit.app
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Upload Dataset
educati...ucce...
27.7KB
Successfully loaded 400 rows × 19 columns from education_career_success.csv
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Statistical Intelligence Engine
Test Recommender
Distribution Analyzer
Hypothesis Builder
Auto-recommend the right statistical test for any two columns.
Variable 1:
Student_ID
Variable 2:
Age
Run Statistical Test

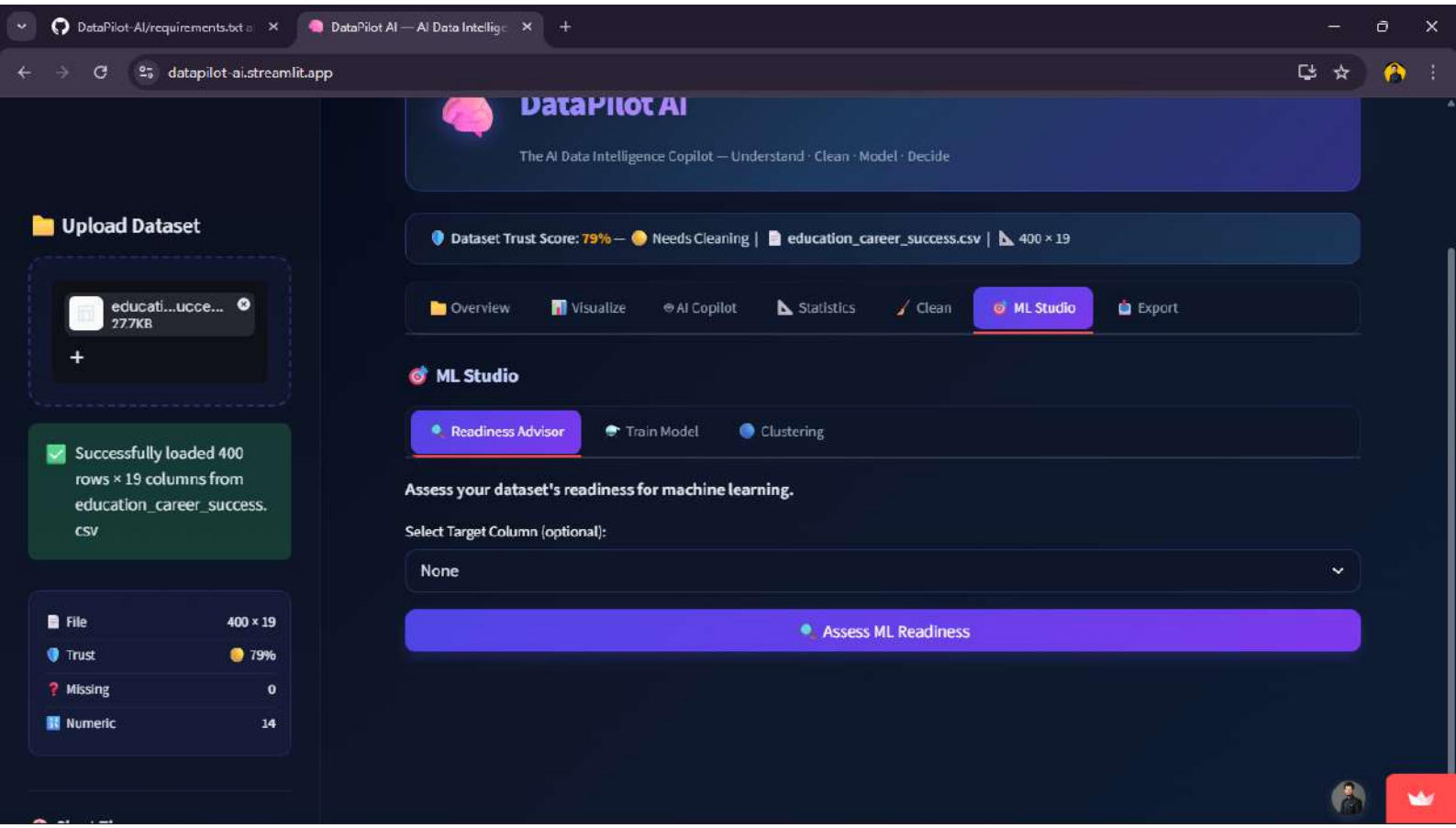

DataPilot-AI/requirements.txt
DataPilot AI — AI Data Intellig
datapilot-ai.streamlit.app
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Upload Dataset
educati...ucce...
27.7KB
Successfully loaded 400 rows × 19 columns from education_career_success.csv
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Data Cleaning Studio
Clean your dataset interactively. Preview changes before applying.
Found 3 area(s) needing attention:
Outliers detected in 3 column(s)
Constant columns: ['Entrepreneurship']
1 Missing Value Handling
No missing values
2 Duplicate Rows
No duplicates


DataPilot-AI/requirements.txt
DataPilot AI — AI Data Intellig
datapilot-ai.streamlit.app
Upload Dataset
educati...ucce...
27.7KB
Successfully loaded 400 rows × 19 columns from education_career_success.csv
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Export & Report Center
Executive Report
Jupyter Notebook
Data Export
Generate a one-click professional analysis report.
Include AI-generated executive summary (requires Groq API)
Add Custom Insights (Optional)
Generate PDF Report
Generate HTML Report
Preview AI Summary

DataPilot AI — AI Data Intellig
Streamlit
datapilot-ai.streamlit.app
Upload Dataset
Successfully loaded 400 rows × 19 columns from education_career_success.csv
Chart Theme
plotly_white
Force App Refresh
DataPilot AI
The AI Data Intelligence Copilot – Understand · Clean · Model · Decide
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Dataset Trust Score
Needs Cleaning
79.5
3 Issue(s) Detected
Possible PII detected: 'Student_ID' (matches 'id')
Possible PII detected: 'Age' (matches 'age')
Possible PII detected: 'Gender' (matches 'gender')
Possible PII detected: 'Internships_Completed' (matches 'ip')
Possible PII detected: 'Entrepreneurship' (matches 'ip')
Recommendations
Review flagged columns for PII before sharing data
Dataset Preview


DataPilot-AI/requirements.txt
DataPilot AI — AI Data Intellig
datapilot-ai.streamlit.app
Upload Dataset
educati...ucce...
27.7KB
Successfully loaded 400 rows × 19 columns from education_career_success.csv
File 400 × 19
Trust 79%
Missing 0
Numeric 14
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Dataset Trust Score
Needs Cleaning
50
79.5
0
10
Completeness 100%
Consistency 100%
Variance 10%
Balance 83%
Uniqueness 100%


DataPilot-AI/requirements.txt
DataPilot AI — AI Data Intellig
datapilot-ai.streamlit.app
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Upload Dataset
educati...ucce...
27.7KB
Successfully loaded 400 rows × 19 columns from education_career_success.csv
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Analysis Type:
Univariate
Univariate Analysis
Column:
Age
Chart:
Histogram
Options
Distribution of Age
Frequency
80
70
60
50
40
30


DataPilot-AI/requirements.txt
DataPilot AI — AI Data Intellig
datapilot-ai.streamlit.app
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Upload Dataset
educati...ucce...
27.7KB
Successfully loaded 400 rows × 19 columns from education_career_success.csv
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
ML Studio
Readiness Advisor
Train Model
Clustering
Assess your dataset's readiness for machine learning.
Select Target Column (optional):
None
Assess ML Readiness

DataPilot AI — AI Data Intellige | DataPilot AI — AI Data Intellige

datapilot-ai.streamlit.app

Upload Dataset

Successfully loaded 400 rows × 19 columns from education_career_succe ss.csv

File	400 × 19
Trust	79%
Missing	0
Numeric	14

Chart Theme

plotly_white

Force App Refresh

Made with ❤ by Nassar Jamath

DataPilot AI v1.0

Dataset Preview

Rows	Columns	Memory	Duplicates
400	19	0.07 MB	0

Data Quality

Quality Score	Completeness	Uniqueness
79.5%	100.0%	100.0%

View details

Column Types

Numeric	Categorical	DateTime	Text	Boolean
14	4	0	1	0
View	View		View	

- Student_ID

Data Preview

First rows | Last rows

	Student_ID	Age	Gender	High_School_GPA	SAT_Score	University_GPA	Field_of_Study	Internships_Completed	Projects_Completed	Certifications	Soft_Skills_Score	Networking_Score	Job_Offers	Starting_Salary	Career_Satisfaction	Years_to_Promotion	Current_Job_Level	Work_Life_Balance	Entrepreneurship
0	S001	22	Male	3.8	1450	3.6	Computer Science	3	7	2	8	7	3	85000	8	2	Mid	7	No
1	S002	24	Female	3.6	1380	3.4	Business	2	5	3	7	6	2	65000	7	3	Mid	6	No
2	S003	21	Male	3.9	1570	3.8	Engineering	4	9	4	9	8	4	120000	9	1	Senior	6	No
3	S004	23	Female	3.5	1300	3.2	Psychology	1	3	1	6	5	1	48000	6	4	Entry	8	No
4	S005	25	Male	3.7	1470	3.5	Medicine	7	6	7	8	7	3	95000	8	2	Mid	5	No
5	S006	22	Female	3.4	1290	3.1	Education	1	4	1	7	6	1	42000	7	5	Entry	8	No
6	S007	24	Male	3.2	1180	2.9	Arts	0	2	0	6	4	0	35000	5	5	Entry	7	No
7	S008	23	Female	3.8	1480	3.7	Law	3	8	3	9	8	3	110000	9	2	Mid	6	No
8	S009	23	Male	3.6	1350	3.3	Finance	2	5	2	7	7	2	72000	7	3	Mid	6	No
9	S010	25	Female	3.9	1550	3.9	Computer Science	4	9	5	10	9	5	140000	10	1	Senior	5	No

Column Information

DataPilot AI — AI Data Intellig

DataPilot AI — AI Data Intellig

datapilot-ai.streamlit.app

Upload Dataset

Successfully loaded 400 rows × 19 columns from education_career_success.csv

File	400 × 19
Trust	78%
Missing	0
Numeric	14

Chart Theme

plotly_white

Force App Refresh

Made with ❤ by Nawaal Zainab

DataPilot AI v1.0

Column Information

Filter by type:

Text × Numeric × Categorical ×

Show columns with missing values only

Column	Data Type	Category	Missing	Missing %	Unique
Gender	str	Categorical	0	0.0%	2
High_School_GPA	float64	Numeric	0	0.0%	8
SAT_Score	int64	Numeric	0	0.0%	35
University_GPA	float64	Numeric	0	0.0%	12
Field_of_Study	str	Categorical	0	0.0%	11
Internships_Completed	int64	Numeric	0	0.0%	5
Projects_Completed	int64	Numeric	0	0.0%	8
Certifications	int64	Numeric	0	0.0%	6
Soft_Skills_Score	int64	Numeric	0	0.0%	6
Networking_Score	int64	Numeric	0	0.0%	8

Descriptive Statistics

Descriptive | Missing Values | Correlations | Outliers | Categorical

Descriptive Statistics

	Count	Mean	Std Dev	Min	Q1	Median	Q3	Max	Range	IQR	Skewness	Kurtosis	Mode
Age	400	23	1.416	21	22	23	24	25	4	2	0	-1.381	21
High_School_GPA	400	3.624	0.188	3.2	3.5	3.6	3.8	3.9	0.7	0.3	0.222	-0.743	3.6
SAT_Score	400	1386.6	114.112	1160	1300	1380	1480	1580	420	180	-0.033	-1.067	1380
University_GPA	400	3.441	0.386	2.8	3.2	3.45	3.7	4	1.2	0.5	-0.129	-0.885	3.5
Internships_Completed	400	2.5	1.063	0	2	3	3	4	4	1	-0.271	-0.743	3
Projects_Completed	400	6.235	1.902	2	5	6	8	9	7	3	-0.278	-0.752	7
Certifications	400	2.623	1.27	0	2	3	3	5	5	1	0.297	-0.404	2
Soft_Skills_Score	400	7.932	1.314	5	7	8	9	10	5	2	-0.051	-0.845	8
Networking_Score	400	6.932	1.118	4	6	7	8	9	5	2	-0.08	-0.82	7
Job_Offers	400	2.74	1.065	0	2	3	4	5	5	2	0.171	-0.651	2

Quick Insights

Age: Distribution is approximately symmetric

High_School_GPA: Distribution is approximately symmetric

SAT_Score: Distribution is approximately symmetric

DataPilot AI — AI Data Intellig

DataPilot AI — AI Data Intellig

datapilot-ai.streamlit.app

Upload Dataset

Successfully loaded 400 rows × 19 columns from education_career_succe ss.csv

File	400 × 19
Trust	78%
Missing	0
Numeric	14

Chart Theme

plotly_white

Force App Refresh

Age: Distribution is approximately symmetric

High_School_GPA: Distribution is approximately symmetric

SAT_Score: Distribution is approximately symmetric

University_GPA: Distribution is approximately symmetric

Internships_Completed: Distribution is approximately symmetric

Projects_Completed: Distribution is approximately symmetric

Certifications: Distribution is approximately symmetric

Soft_Skills_Score: Distribution is approximately symmetric

Networking_Score: Distribution is approximately symmetric

Job_Offers: Distribution is approximately symmetric

Starting_Salary: Distribution is approximately symmetric

Career_Satisfaction: Distribution is approximately symmetric

Years_to_Promotion: Distribution is approximately symmetric

Work_Life_Balance: Distribution is approximately symmetric

Correlation Matrix

Correlation Matrix

	Age	High_School_GPA	SAT_Score	University_GPA	Internships_Completed	Projects_Completed	Certifications	Soft_Skills_Score	Networking_Score	Job_Offers	Starting_Salary	Career_Satisfaction	Years_to_Promotion	Work_Life_Balance
Age	1.00	0.19	0.16	0.18	0.24	0.20	0.22	0.16	0.17	0.24	0.20	0.24	-0.17	-0.12
High_School_GPA	0.19	1.00	0.98	0.98	0.95	0.98	0.94	0.97	0.97	0.97	0.95	0.98	-0.98	-0.80
SAT_Score	0.16	0.98	1.00	0.98	0.94	0.95	0.94	0.96	0.96	0.97	0.95	0.97	-0.98	-0.80
University_GPA	0.18	0.98	0.98	1.00	0.96	0.98	0.94	0.97	0.97	0.97	0.96	0.97	-0.98	-0.81
Internships_Completed	0.24	0.95	0.94	0.96	1.00	0.96	0.97	0.93	0.94	0.97	0.94	0.98	-0.94	-0.86
Projects_Completed	0.20	0.98	0.95	0.98	0.96	1.00	0.94	0.94	0.94	0.97	0.94	0.98	-0.93	-0.79
Certifications	0.22	0.94	0.94	0.94	0.97	0.94	1.00	0.94	0.94	0.97	0.97	0.98	-0.93	-0.83
Soft_Skills_Score	0.16	0.97	0.96	0.97	0.93	0.94	0.94	1.00	1.00	0.96	0.95	0.97	-0.94	-0.83
Networking_Score	0.17	0.97	0.96	0.97	0.94	0.94	0.94	1.00	1.00	0.96	0.95	0.97	-0.94	-0.83
Job_Offers	0.24	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.96	1.00	0.97	0.99	-0.94	-0.82
Starting_Salary	0.20	0.95	0.95	0.96	0.94	0.94	0.97	0.95	0.95	0.97	1.00	0.97	-0.93	-0.88
Career_Satisfaction	0.24	0.98	0.97	0.97	0.98	0.98	0.98	0.97	0.97	0.99	0.97	1.00	-0.94	-0.81
Years_to_Promotion	-0.17	-0.98	-0.98	-0.98	-0.94	-0.93	-0.93	-0.94	-0.94	-0.94	-0.93	-0.94	1.00	0.83
Work_Life_Balance	-0.12	-0.80	-0.80	-0.81	-0.86	-0.79	-0.83	-0.83	-0.83	-0.82	-0.88	-0.81	0.83	1.00

Download Chart

PNG JPEG SVG HTML CSV Raw

DataPilot AI — AI Data Intellig

DataPilot AI — AI Data Intellig

datapilot-ai.streamlit.app

Upload Dataset

educ
27.7K

File 400 × 19
Trust 79%
Missing 0
Numeric 14

Chart Theme

plotly_wh...

Force App Refresh

Made with by Mansoor Sarookh
DataPilot AI v2.0

Filter by type:

Text × Numeric ×

Show columns with missing values only

Column	Data Type	Category	Missing	Missing %	Unique
Student_ID	str	Text	0	0.0%	400
Age	int64	Numeric	0	0.0%	5
High_School_GPA	float64	Numeric	0	0.0%	8
SAT_Score	int64	Numeric	0	0.0%	35
University_GPA	float64	Numeric	0	0.0%	13
Internships_Completed	int64	Numeric	0	0.0%	5
Projects_Completed	int64	Numeric	0	0.0%	8
Certifications	int64	Numeric	0	0.0%	6
Soft_Skills_Score	int64	Numeric	0	0.0%	6
Networking_Score	int64	Numeric	0	0.0%	6

Descriptive Statistics

Descriptive | Missing Values | Correlations | Outliers | Categorical

Descriptive Statistics

	Count	Mean	Std Dev	Min	Q1	Median	Q3	Max	Range	IQR	Skewness	Kurtosis	Mode
Age	400	23	1.416	21	22	23	24	25	4	2	0	-1.301	21
High_School_GPA	400	3.624	0.188	3.2	3.5	3.6	3.8	3.9	0.7	0.3	-0.222	-0.783	3.6
SAT_Score	400	1388.6	114.312	1160	1300	1380	1480	1580	420	180	-0.033	-1.067	1380
University_GPA	400	3.441	0.286	2.8	3.2	3.45	3.7	4	1.2	0.5	-0.129	-0.805	

DataPilot AI — AI Data Intellig

DataPilot AI — AI Data Intellig

datapilot-ai.streamlit.app

Career_Satisfaction: Distribution is approximately symmetric

Years_to_Promotion: Distribution is approximately symmetric

Work_Life_Balance: Distribution is approximately symmetric

Upload Dataset

edu
27.7K

File 400 × 19

Trust 79%

Missing 0

Numeric 14

Chart Theme

plotly_wh...

Force App Refresh

Made with by Mansoor Sarookh

DataPilot AI v2.0

Correlation Matrix

Correlation Matrix

	Age	High_School_GPA	SAT_Score	University_GPA	Internships_Completed	Projects_Completed	Certifications	Soft_Skills_Score	Networking_Score	Job_Offers	Starting_Salary	Career_Satisfaction	Years_to_Promotion	Work_Life_Balance
Age	1.00	0.19	0.18	0.18	0.24	0.20	0.22	0.18	0.17	0.24	0.20	0.24	-0.17	-0.13
High_School_GPA	0.19	1.00	0.98	0.98	0.95	0.96	0.94	0.97	0.97	0.97	0.95	0.96	-0.96	-0.80
SAT_Score	0.18	0.98	1.00	0.98	0.94	0.95	0.94	0.96	0.96	0.97	0.95	0.97	-0.96	-0.80
University_GPA	0.18	0.98	0.98	1.00	0.95	0.96	0.94	0.97	0.97	0.97	0.96	0.97	-0.96	-0.83
Internships_Completed	0.24	0.95	0.94	0.95	1.00	0.96	0.97	0.93	0.94	0.97	0.94	0.96	-0.94	-0.80
Projects_Completed	0.20	0.96	0.95	0.96	0.96	1.00	0.94	0.94	0.94	0.97	0.94	0.96	-0.93	-0.79
Certifications	0.22	0.94	0.94	0.94	0.97	0.94	1.00	0.94	0.94	0.97	0.97	0.96	-0.93	-0.82
Soft_Skills_Score	0.18	0.97	0.96	0.97	0.93	0.94	0.94	1.00	1.00	0.96	0.95	0.97	-0.94	-0.83
Networking_Score	0.17	0.97	0.96	0.97	0.94	0.94	0.94	1.00	1.00	0.96	0.95	0.97	-0.94	-0.83
Job_Offers	0.24	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.96	1.00	0.97	0.99	-0.94	-0.82
Starting_Salary	0.20	0.95	0.95	0.96	0.94	0.94	0.97	0.95	0.95	0.97	1.00	0.97	-0.93	-0.85
Career_Satisfaction	0.24	0.96	0.97	0.97	0.96	0.96	0.96	0.97	0.97	0.99	0.97	1.00	-0.94	-0.81
Years_to_Promotion	-0.17	-0.96	-0.96	-0.96	-0.94	-0.93	-0.93	-0.94	-0.94	-0.94	-0.93	-0.94	1.00	0.83
Work_Life_Balance	-0.13	-0.80	-0.80	-0.83	-0.80	-0.79	-0.82	-0.83	-0.83	-0.82	-0.85	-0.81	0.83	1.00

1 0.5 0 −0.5

Download Chart

PNG JPEG SVG HTML Hi-Res

DataPilot AI — AI Data Intellig

datapilot-ai.streamlit.app

Internships_Completed	int64	Numeric	0	0.0%	5
Projects_Completed	int64	Numeric	0	0.0%	8
Certifications	int64	Numeric	0	0.0%	6
Soft_Skills_Score	int64	Numeric	0	0.0%	6
Networking_Score	int64	Numeric	0	0.0%	6

Upload Dataset

educ
27.7K

File 400 × 19
Trust 79%
Missing 0
Numeric 14

Chart Theme

plotly_wh...

Force App Refresh

Made with by Mansoor Sarookh
DataPilot AI v2.0

Descriptive Statistics

Descriptive | Missing Values | Correlations | Outliers | Categorical

Correlation Insights

Found 78 highly correlated pairs (|r| ≥ 0.7):

Variable 1	Variable 2	Correlation	Strength
High_School_GPA	SAT_Score	0.9824	Strong Positive
High_School_GPA	University_GPA	0.9832	Strong Positive
High_School_GPA	Internships_Completed	0.9499	Strong Positive
High_School_GPA	Projects_Completed	0.9598	Strong Positive
High_School_GPA	Certifications	0.9368	Strong Positive
High_School_GPA	Soft_Skills_Score	0.9655	Strong Positive
High_School_GPA	Networking_Score	0.9667	Strong Positive
High_School_GPA	Job_Offers	0.9664	Strong Positive
High_School_GPA	Starting_Salary	0.9505	Strong Positive
High_School_GPA	Career_Satisfaction	0.9638	Strong Positive

DataPilot AI — AI Data Intellig

DataPilot AI — AI Data Intellig

datapilot-ai.streamlit.app

Soft_Skills_Score	400	0.932	1.114	5	7	8	9	10	5	2	-0.061	-0.945	8
Networking_Score	400	6.931	1.118	4	6	7	8	9	5	3	-0.08	-0.17	7
Job_Offers	400	2.74	1.155	0	2	3	4	5	5	2	0.171	-0.881	2

Upload Dataset

education_ca... 377KB

Successfully loaded 400 rows × 19 columns from education_career_success.csv

File	400 × 19
Trust	78%
Missing	0
Numeric	14

Chart Theme

plotly_white

Force App Refresh

Made with ❤ by Nawaal Zainab

DataPilot AI v1.0

Quick Insights

Age: Distribution is approximately symmetric

High_School_GPA: Distribution is approximately symmetric

SAT_Score: Distribution is approximately symmetric

University_GPA: Distribution is approximately symmetric

Internships_Completed: Distribution is approximately symmetric

Projects_Completed: Distribution is approximately symmetric

Certifications: Distribution is approximately symmetric

Soft_Skills_Score: Distribution is approximately symmetric

Networking_Score: Distribution is approximately symmetric

Job_Offers: Distribution is approximately symmetric

Starting_Salary: Distribution is approximately symmetric

Career_Satisfaction: Distribution is approximately symmetric

Years_to_Promotion: Distribution is approximately symmetric

Work_Life_Balance: Distribution is approximately symmetric

Correlation Matrix

Correlation Matrix

	Age	High_School_GPA	SAT_Score	University_GPA	Internships_Completed	Projects_Completed	Certifications	Soft_Skills_Score	Networking_Score	Job_Offers	Starting_Salary	Career_Satisfaction	Years_to_Promotion	Work_Life_Balance
Age	1.00	0.19	0.16	0.18	0.24	0.20	0.22	0.18	0.17	0.24	0.20	0.24	-0.17	-0.12
High_School_GPA	0.19	1.00	0.98	0.98	0.95	0.98	0.94	0.97	0.97	0.97	0.95	0.98	-0.98	-0.80
SAT_Score	0.16	0.98	1.00	0.98	0.94	0.95	0.94	0.96	0.96	0.97	0.95	0.97	-0.98	-0.80
University_GPA	0.18	0.98	0.98	1.00	0.95	0.98	0.94	0.97	0.97	0.97	0.96	0.97	-0.98	-0.83
Internships_Completed	0.24	0.95	0.94	0.95	1.00	0.98	0.97	0.93	0.94	0.97	0.94	0.98	-0.94	-0.88
Projects_Completed	0.20	0.98	0.95	0.98	0.98	1.00	0.94	0.94	0.94	0.97	0.94	0.98	-0.93	-0.79
Certifications	0.22	0.94	0.94	0.94	0.97	0.94	1.00	0.94	0.94	0.97	0.97	0.98	-0.93	-0.82
Soft_Skills_Score	0.18	0.97	0.96	0.97	0.93	0.94	0.94	1.00	1.00	0.96	0.95	0.97	-0.94	-0.83
Networking_Score	0.17	0.97	0.96	0.97	0.94	0.94	0.94	1.00	1.00	0.96	0.95	0.97	-0.94	-0.83
Job_Offers	0.24	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.96	1.00	0.97	0.99	-0.94	-0.83
Starting_Salary	0.20	0.95	0.95	0.96	0.94	0.94	0.97	0.95	0.95	0.97	1.00	0.97	-0.95	-0.84
Career_Satisfaction	0.24	0.98	0.97	0.97	0.98	0.98	0.98	0.97	0.97	0.99	0.97	1.00	-0.94	-0.81
Years_to_Promotion	-0.17	-0.98	-0.98	-0.98	-0.94	-0.93	-0.93	-0.94	-0.94	-0.94	-0.95	-0.94	1.00	0.83
Work_Life_Balance	-0.12	-0.80	-0.80	-0.83	-0.88	-0.79	-0.82	-0.83	-0.83	-0.83	-0.84	-0.81	0.83	1.00

1 0.5 0 −0.5

DataPilot AI — AI Data Intellig

DataPilot AI — AI Data Intellig

datapilot-ai.streamlit.app

Upload Dataset

edu
27.7K

File 400 × 19
Trust 79%
Missing 0
Numeric 14

Chart Theme

plotly_wh...

Force App Refresh

Made with by Mansoor Sarookh
DataPilot AI v2.0

Descriptive Statistics

Descriptive | Missing Values | Correlations | Outliers | Categorical

Descriptive Statistics

	Count	Mean	Std Dev	Min	Q1	Median	Q3	Max	Range	IQR	Skewness	Kurtosis	Mode
Age	400	23	1.416	21	22	23	24	25	4	2	0	-1.301	21
High_School_GPA	400	3.624	0.188	3.2	3.5	3.6	3.8	3.9	0.7	0.3	-0.222	-0.783	3.6
SAT_Score	400	1388.6	114.312	1160	1300	1380	1480	1580	420	180	0.033	-1.067	1380
University_GPA	400	3.441	0.286	2.8	3.2	3.45	3.7	4	1.2	0.5	-0.129	-0.805	3.5
Internships_Completed	400	2.5	1.062	0	2	3	3	4	4	1	-0.271	-0.742	3
Projects_Completed	400	6.235	1.902	2	5	6	8	9	7	3	-0.278	-0.752	7
Certifications	400	2.625	1.27	0	2	3	3	5	5	1	0.297	-0.424	3
Soft_Skills_Score	400	7.932	1.314	5	7	8	9	10	5	2	-0.061	-0.845	8
Networking_Score	400	6.932	1.318	4	6	7	8	9	5	2	-0.08	-0.82	7
Job_Offers	400	2.74	1.355	0	2	3	4	5	5	2	0.171	-0.851	2

Quick Insights

Age: Distribution is approximately symmetric

High_School_GPA: Distribution is approximately symmetric

SAT_Score: Distribution is approximately symmetric

University_GPA: Distribution is approximately symmetric

Internships_Completed: Distribution is approximately symmetric

Projects_Completed: Distribution is approximately symmetric

Certifications: Distribution is approximately symmetric

DataPilot AI — AI Data Intellig | DataPilot AI — AI Data Intellig

datapilot-ai.streamlit.app

Upload Dataset

educ 27.7K

File 400 × 19
Trust 79%
Missing 0
Numeric 14

Chart Theme

plotly_wh...

Force App Refresh

Made with by Mansoor Sarookh
DataPilot AI v2.0

Student_ID	str	Text	0	0.0%	400
Age	int64	Numeric	0	0.0%	5
High_School_GPA	float64	Numeric	0	0.0%	8
SAT_Score	int64	Numeric	0	0.0%	35
University_GPA	float64	Numeric	0	0.0%	13
Internships_Completed	int64	Numeric	0	0.0%	5
Projects_Completed	int64	Numeric	0	0.0%	8
Certifications	int64	Numeric	0	0.0%	6
Soft_Skills_Score	int64	Numeric	0	0.0%	6
Networking_Score	int64	Numeric	0	0.0%	6

Descriptive Statistics

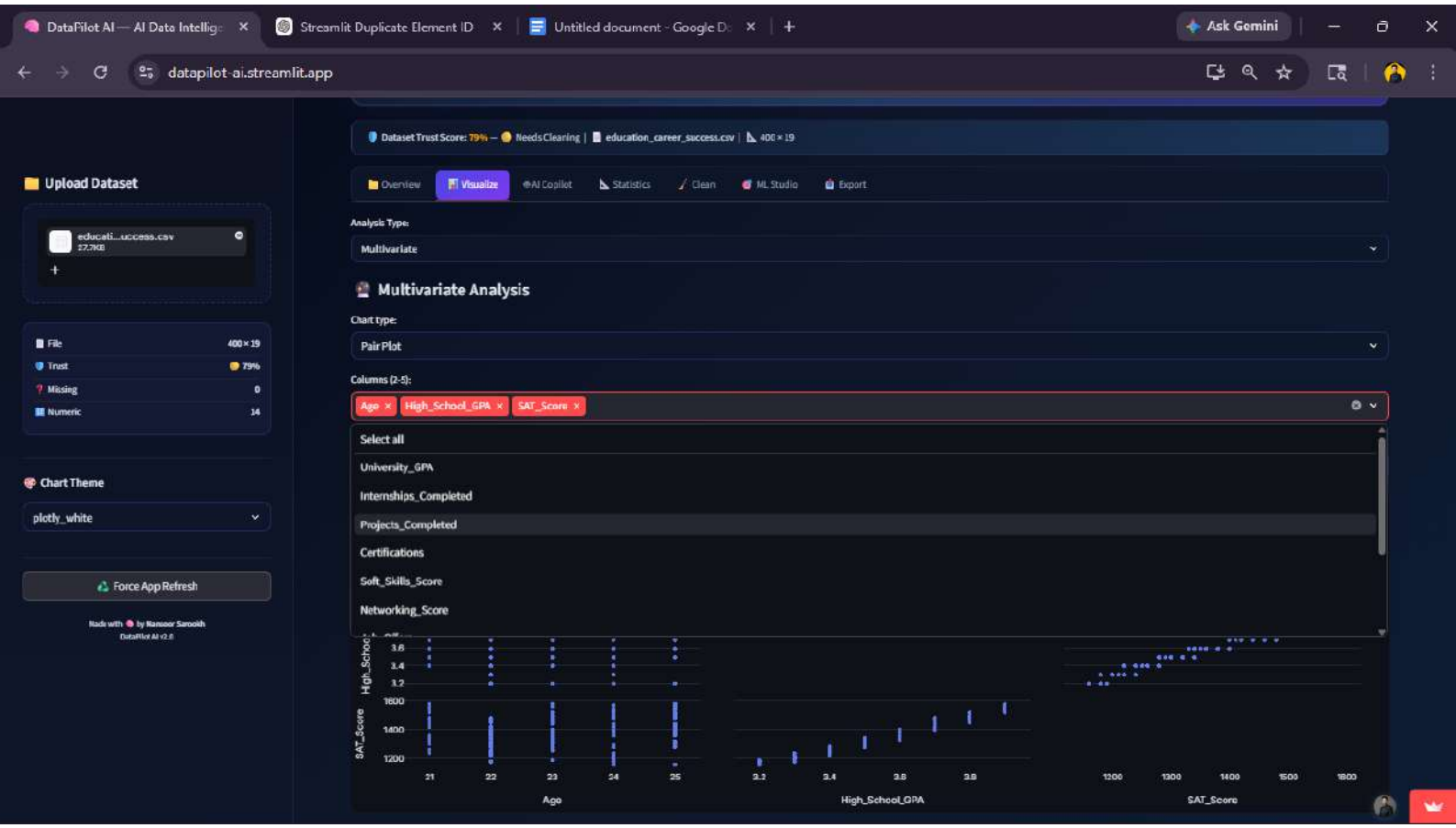
Descriptive | Missing Values | Correlations | Outliers | Categorical

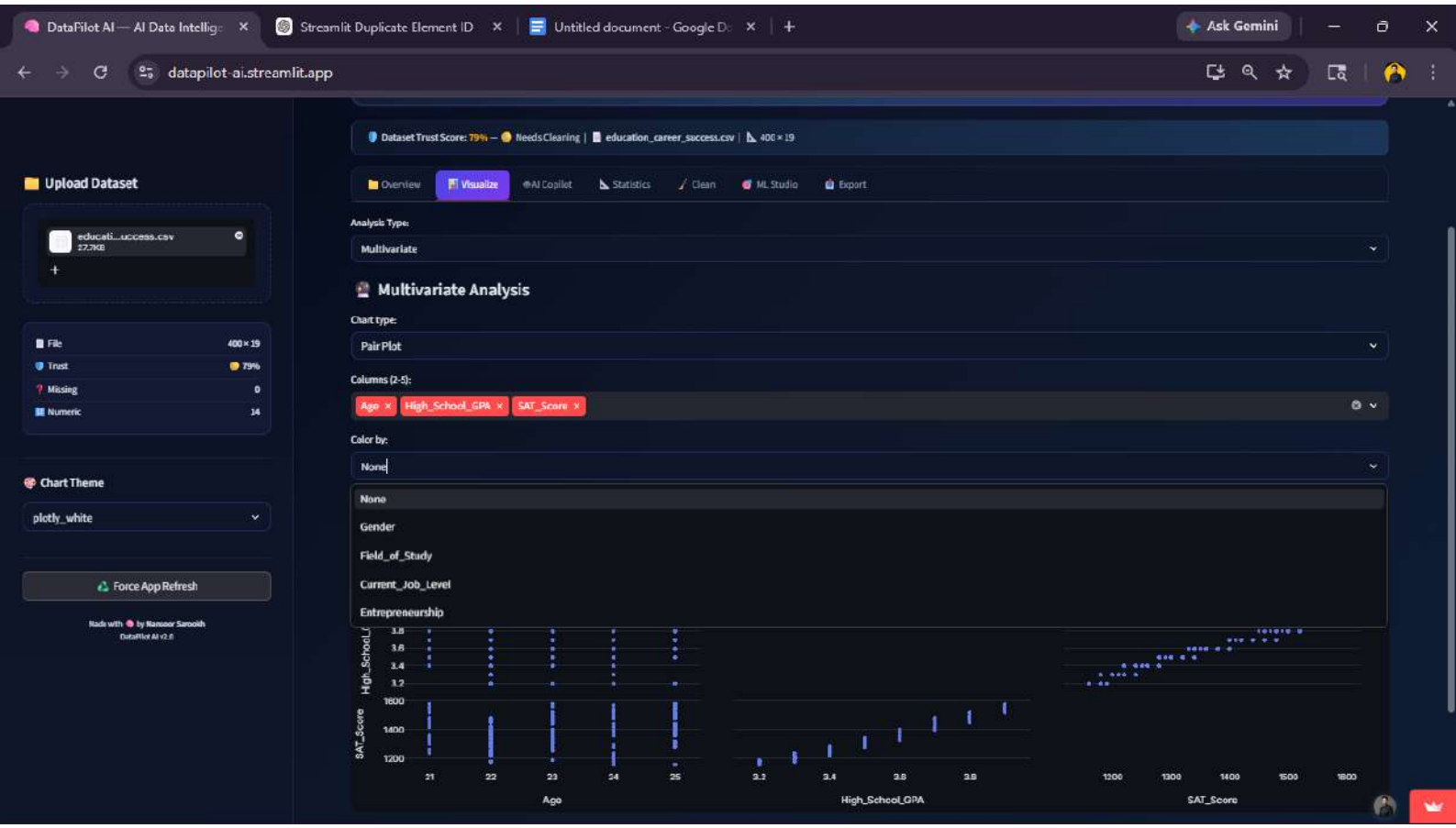
Outlier Detection (IQR Method)

Column	Outliers	Percentage
Internships_Completed	10	2.50%
Certifications	60	15.00%
Work_Life_Balance	1	0.25%

Correlation Matrix

Correlation Matrix


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
educati...uccess.csv
27.7KB
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_white
Force App Refresh
Made with by Nansoor Sarookh
DataPilot AI v2.0
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Analysis Type:
Multivariate
Multivariate Analysis
Chart type:
Pair Plot
Columns (2-5):
Age
High_School_GPA
SAT_Score
Select all
University_GPA
Internships_Completed
Projects_Completed
Certifications
Soft_Skills_Score
Networking_Score
SAT_Score
Age
High_School_GPA
SAT_Score


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
educati...uccess.csv
27.7KB
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_white
Force App Refresh
Made with by Nansoor Sarookh
DataPilot AI v2.0
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Analysis Type:
Multivariate
Multivariate Analysis
Chart type:
Pair Plot
Columns (2-5):
Age
High_School_GPA
SAT_Score
Color by:
None
None
Gender
Field_of_Study
Current_Job_Level
Entrepreneurship
SAT_Score
Age
High_School_GPA
SAT_Score

DataPilot AI — AI Data Intellig | Streamlit Duplicate Element ID | Untitled document - Google D | Ask Gemini

datapilot-ai.streamlit.app

What cleaning steps do you recommend? | Prepare ML-ready feature suggestions | Are there any anomalies or outliers?

Upload Dataset

edu 27.7

File 400 × 19
Trust 79%
Missing 0
Numeric 14

Chart Theme

plotly_w...

Force App Refresh

Made with by Nanoor Sarokh DataPilot AI v2.0

tell me why not this dataset is perfect for the ml.

Insight: The dataset has 400 rows and 19 columns, with no missing values and no duplicates. The mean age of the students is 23.0, and the mean university GPA is 3.441.

Risk: The dataset may not be perfect for ML due to potential biases in the data. For example, the mean SAT score is 1388.6, which is relatively high, indicating that the dataset may be biased towards students from higher socio-economic backgrounds.

Action: To improve the dataset, consider collecting more data from diverse sources to reduce potential biases. Additionally, explore feature engineering techniques to create new features that can help improve model performance. Consider applying statistical tests, such as correlation analysis, to identify relationships between variables.

What trends exist in this dataset?

Insight:

1. The mean number of internships completed is 2.5, and the mean number of projects completed is 6.235, indicating that students have more project experience than internship experience.
2. The mean soft skills score (7.933) is higher than the mean networking score (6.933), suggesting that students may have stronger soft skills than networking skills.
3. The mean starting salary is not provided, but the sample data shows a range of starting salaries from 48,000 to 120,000, indicating a potential relationship between university GPA, internships, and starting salary.

Risk: The dataset may not capture the full range of experiences and outcomes for all students, potentially leading to biased trends.

Action: Apply correlation analysis and visualization techniques (e.g., scatter plots, heatmaps) to explore relationships between variables, such as university GPA, internships, and starting salary, to identify more trends and patterns in the data.

Ask anything about your data... (e.g. 'Why is revenue dropping?')

Clear Chat History

Dataset Context Summary

DataPilot AI — AI Data Intellig | Streamlit Duplicate Element ID | Untitled document - Google D | Ask Gemini

datapilot-ai.streamlit.app

DataPilot AI

The AI Data Intelligence Copilot — Understand · Clean · Model · Decide

Upload Dataset

edu 27.7

+

File 400 × 19
Trust 79%
Missing 0
Numeric 14

Chart Theme

plotly_w...

Force App Refresh

Made with by Nanoor Sarokh
DataPilot AI v2.0

Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19

Overview | Visualize | AI Copilot | Statistics | Clean | ML Studio | Export

Statistical Intelligence Engine

Test Recommender | Distribution Analyzer | Hypothesis Builder

Auto-recommend the right statistical test for any two columns.

Variable 1: Gender

Variable 2: Internships_Completed

Run Statistical Test

Test Selected: `Mann-Whitney U Test`

Test Statistic	P-Value	Significant?
24983.0	0.0000	✅ Yes

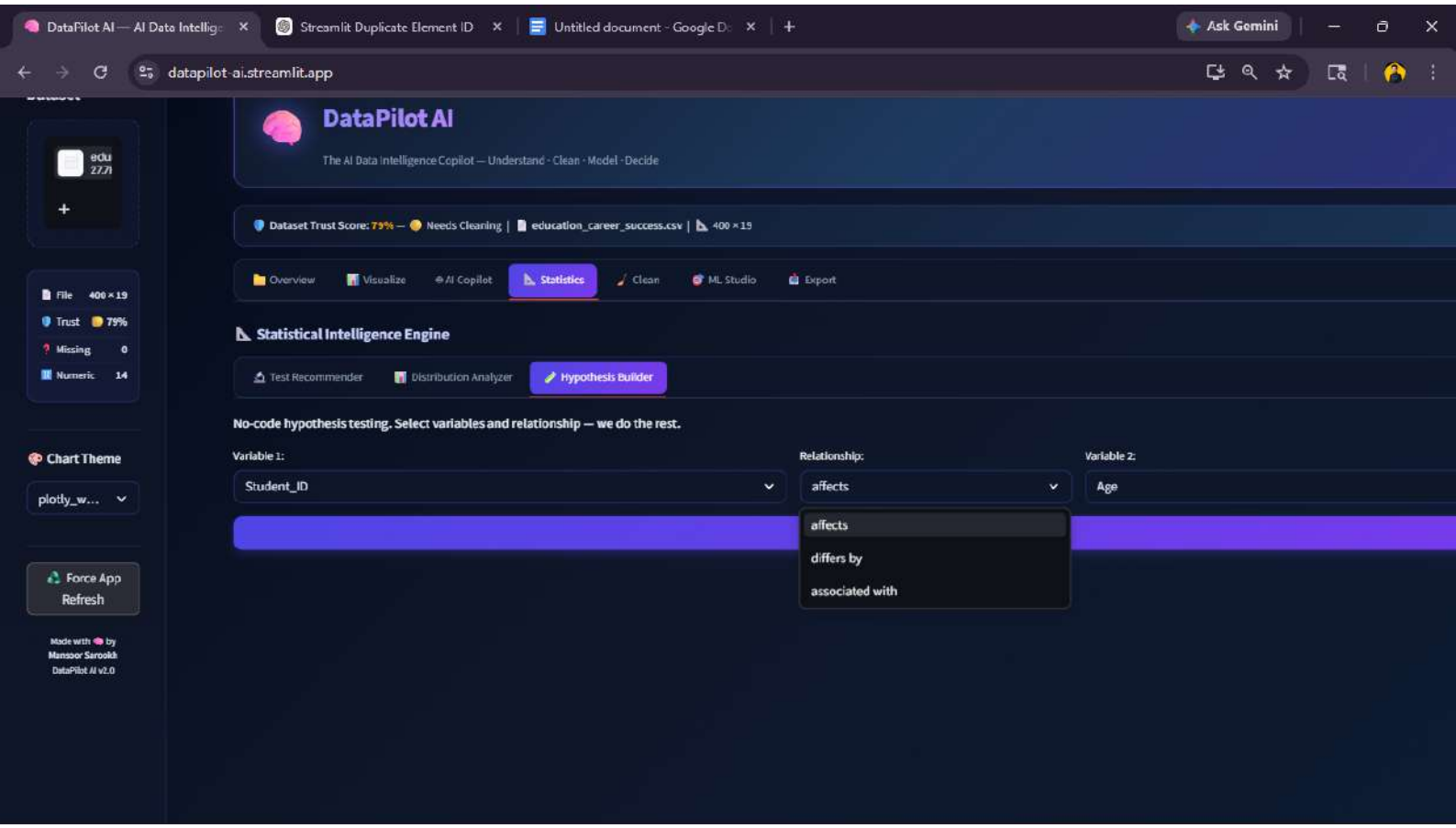
Assumptions Check

normality_ok: False

n_groups: 2

Interpretation: Significant difference in 'Internships_Completed' across 'Gender' groups (Mann-Whitney U Test: stat=24983.000, p=<0.001).

Recommendation: Consider using this grouping as a feature in your ML model.


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 15
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_w...
Force App Refresh
Made with by Mansoor Sarookh DataPilot AI v2.0
Statistical Intelligence Engine
Test Recommender
Distribution Analyzer
Hypothesis Builder
No-code hypothesis testing. Select variables and relationship — we do the rest.
Variable 1:
Student_ID
Relationship:
affects
Variable 2:
Age
affects
differs by
associated with

DataPilot AI — AI Data Intellig

DataPilot AI — AI Data Intellig

datapilot-ai.streamlit.app

Descriptive | Missing Values | Correlations | Outliers | Categorical

Upload Dataset

educ 27.7K

File 400 × 19

Trust 79%

Missing 0

Numeric 14

Chart Theme

plotly_wh...

Force App Refresh

Made with by Mansoor Sarookh

DataPilot AI v2.0

Categorical Columns

Age (5 categories)

Value	Count	Percentage
22	80	20
24	80	20
21	80	20
23	80	20
25	80	20

Gender (2 categories)

Value	Count	Percentage
Male	200	50
Female	200	50

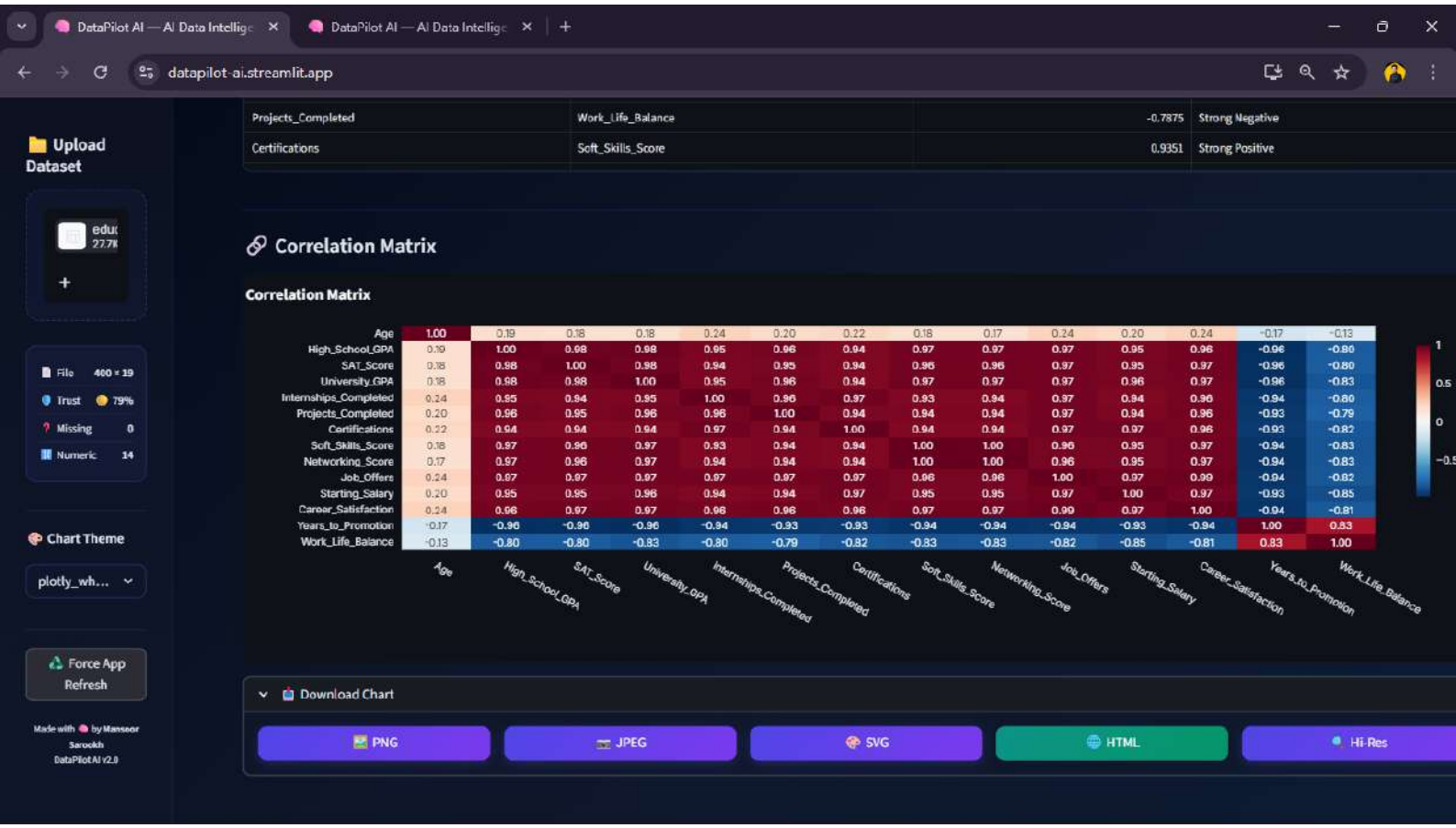
High_School_GPA (8 categories)

University_GPA (13 categories)

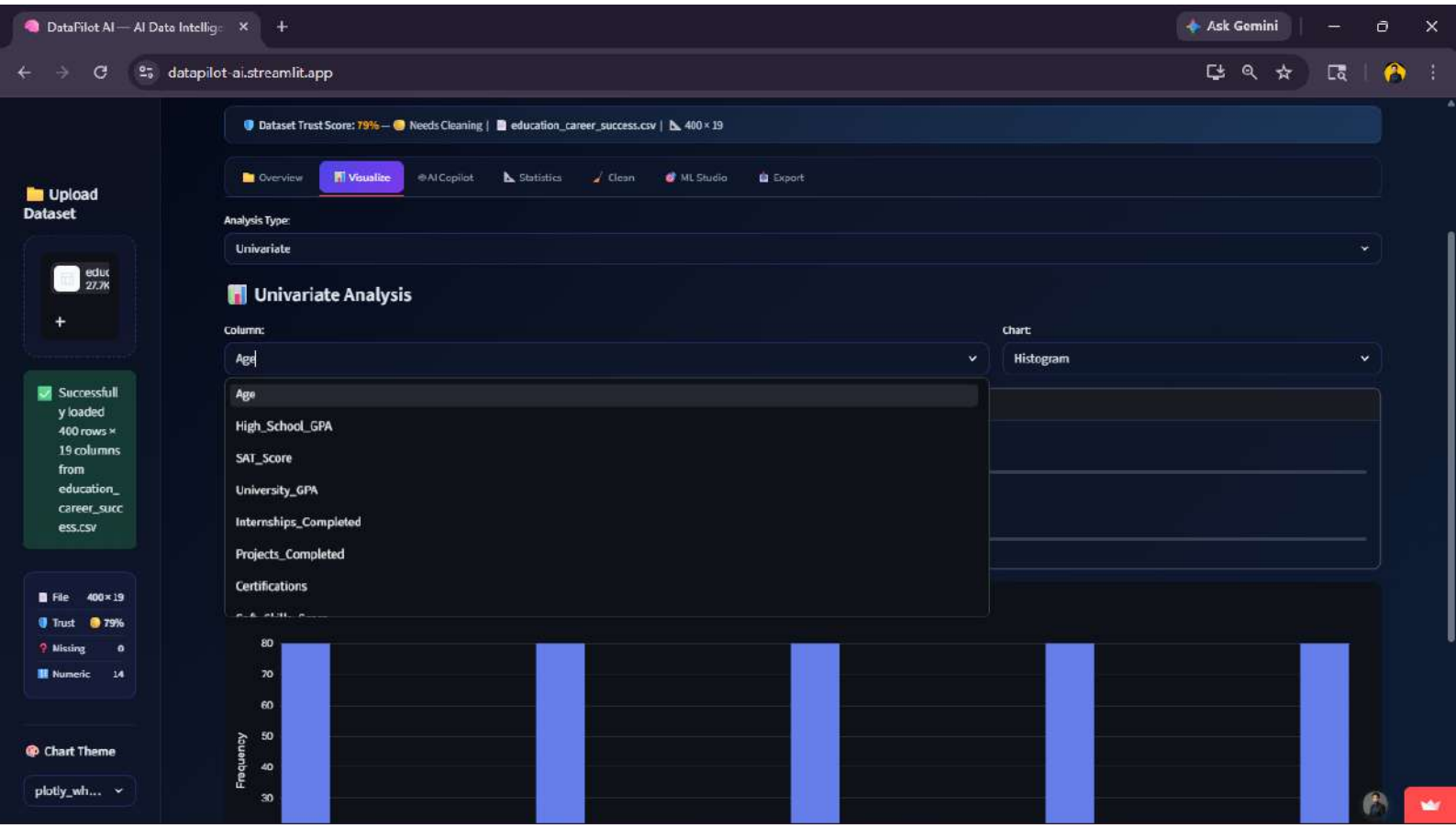
Field_of_Study (11 categories)

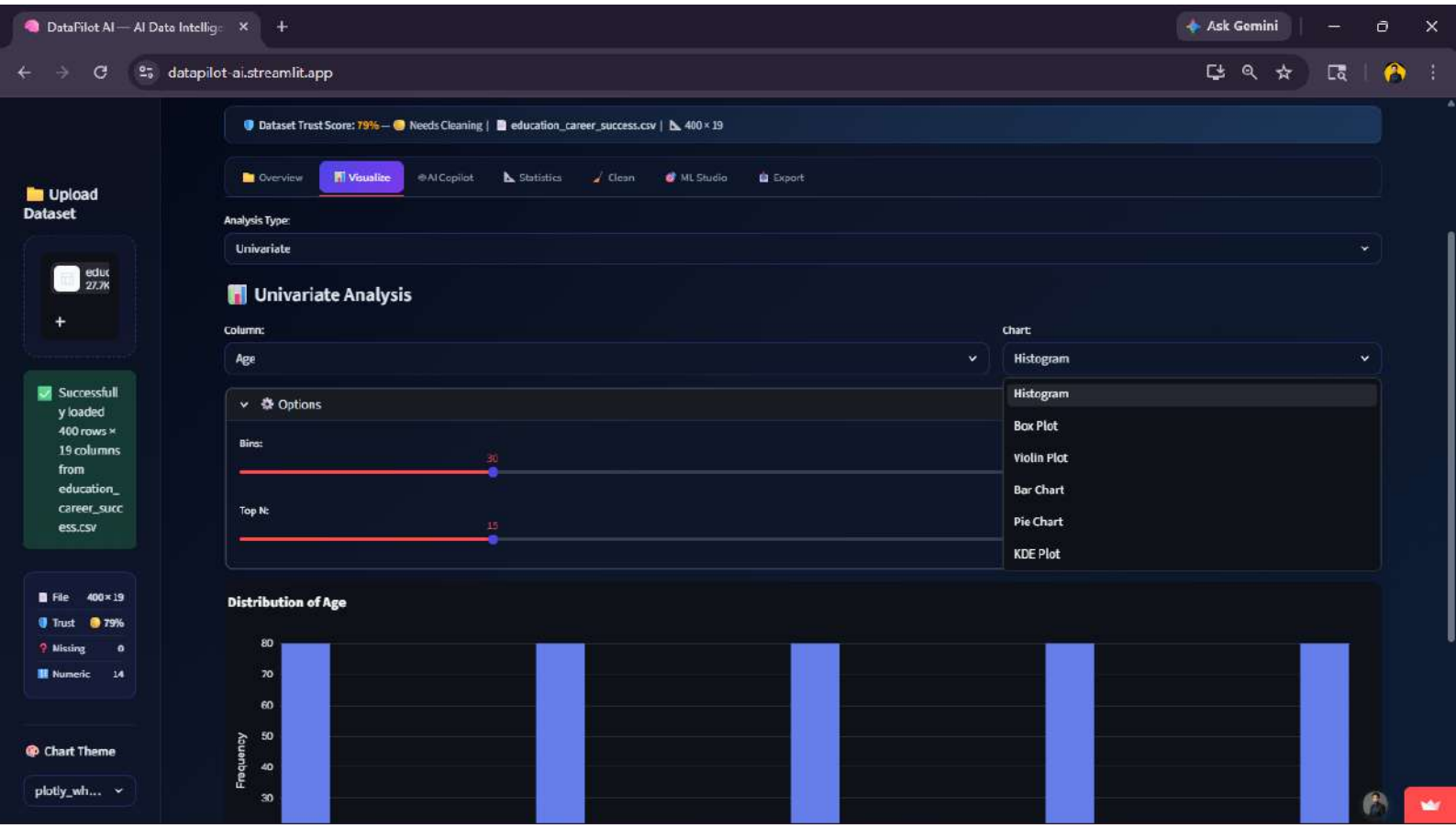
Internships_Completed (5 categories)

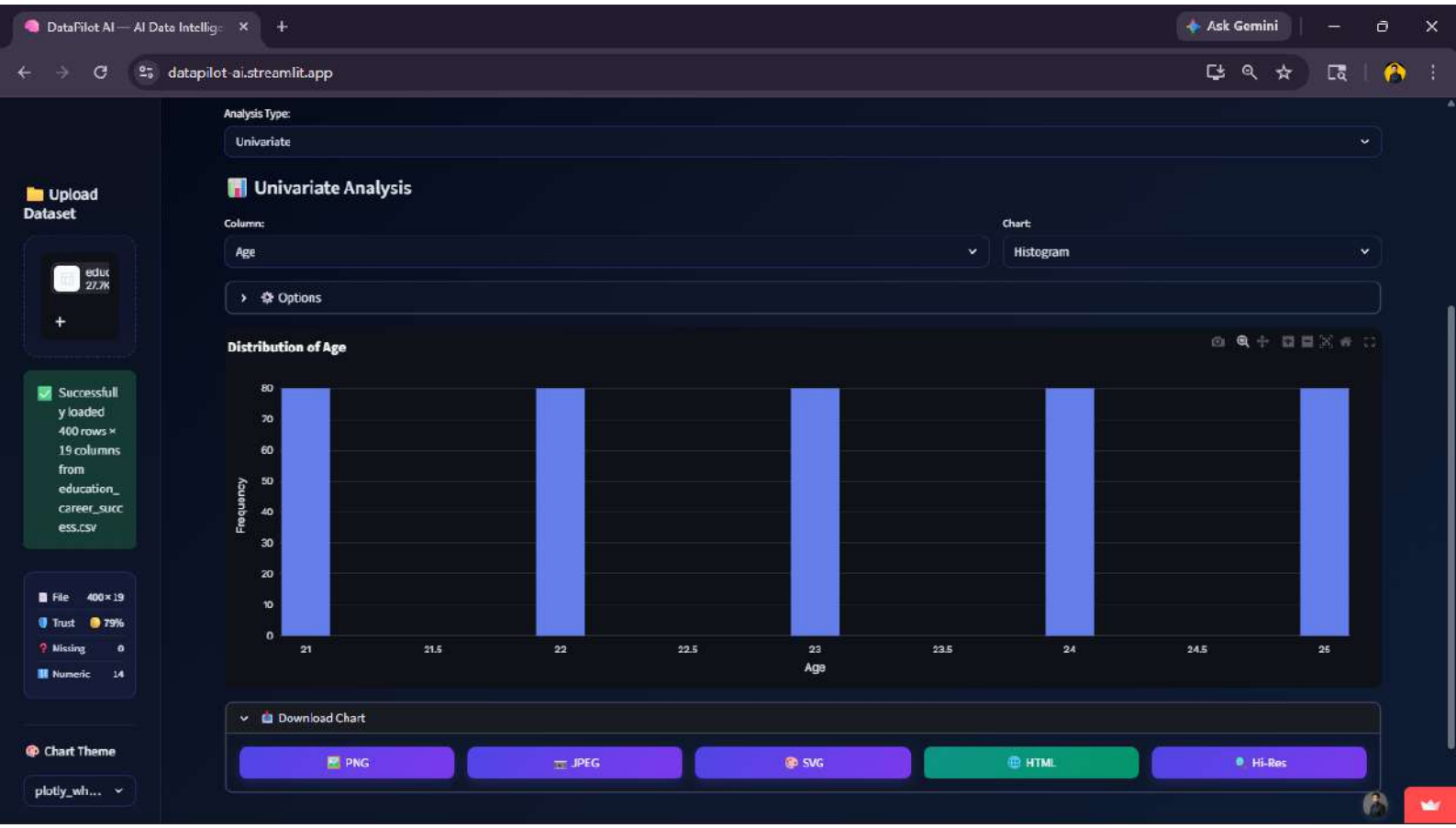
Projects_Completed (8 categories)

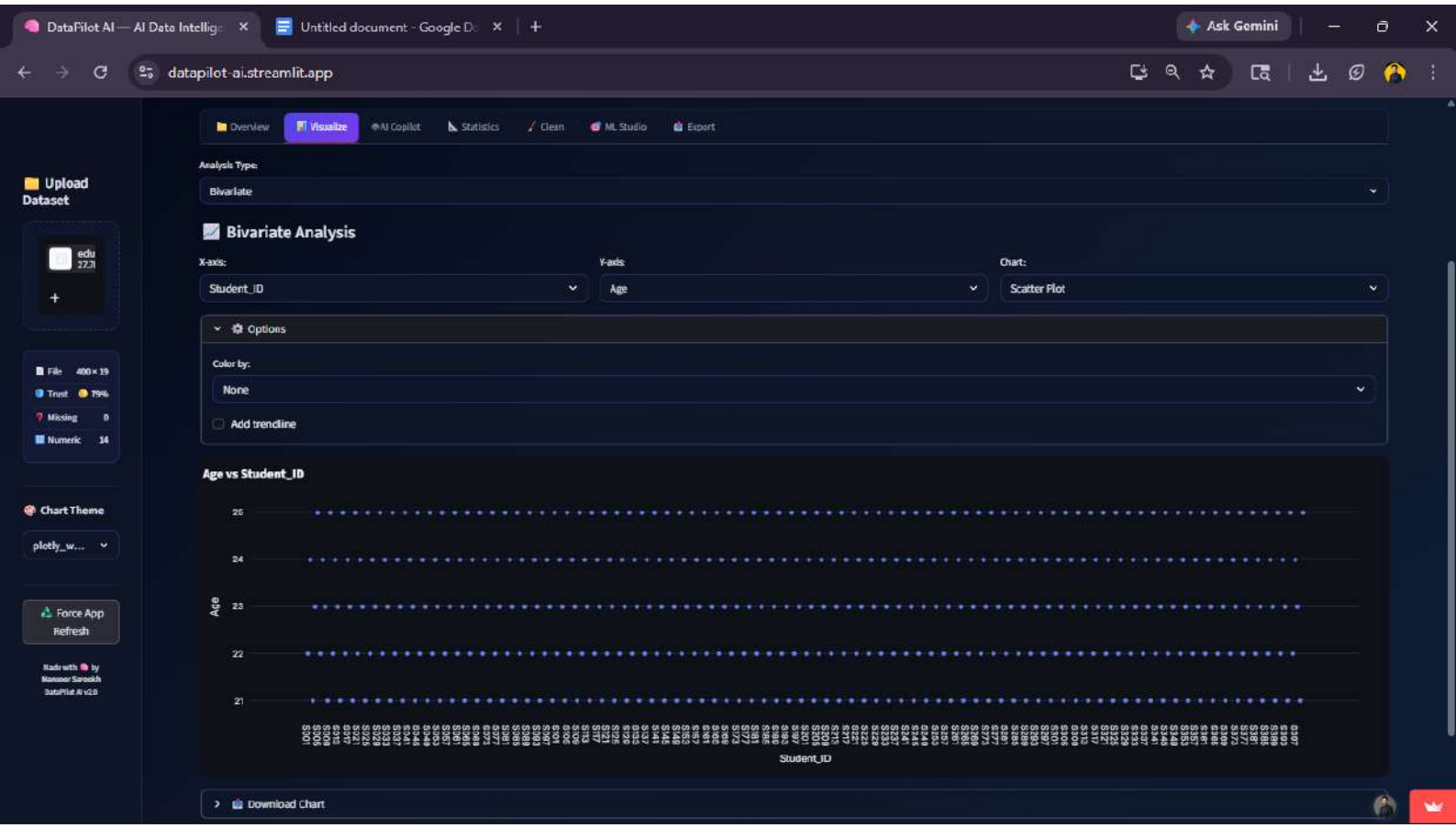


	Age	High_School_GPA	SAT_Score	University_GPA	Internships_Completed	Projects_Completed	Certifications	Soft_Skills_Score	Networking_Score	Job_Offers	Starting_Salary	Career_Satisfaction	Years_to_Promotion	Work_Life_Balance
Age	1.00	0.19	0.18	0.18	0.24	0.20	0.22	0.18	0.17	0.24	0.20	0.24	-0.17	-0.13
High_School_GPA	0.19	1.00	0.98	0.98	0.95	0.96	0.94	0.97	0.97	0.97	0.95	0.96	-0.96	-0.80
SAT_Score	0.18	0.98	1.00	0.98	0.94	0.95	0.94	0.96	0.96	0.97	0.95	0.97	-0.96	-0.80
University_GPA	0.18	0.98	0.98	1.00	0.95	0.96	0.94	0.97	0.97	0.97	0.96	0.97	-0.96	-0.83
Internships_Completed	0.24	0.95	0.94	0.95	1.00	0.96	0.97	0.93	0.94	0.97	0.94	0.96	-0.94	-0.80
Projects_Completed	0.20	0.96	0.95	0.96	0.96	1.00	0.94	0.94	0.94	0.97	0.94	0.96	-0.93	-0.79
Certifications	0.22	0.94	0.94	0.94	0.97	0.94	1.00	0.94	0.94	0.97	0.97	0.96	-0.93	-0.82
Soft_Skills_Score	0.18	0.97	0.96	0.97	0.93	0.94	0.94	1.00	1.00	0.96	0.95	0.97	-0.94	-0.83
Networking_Score	0.17	0.97	0.96	0.97	0.94	0.94	0.94	1.00	1.00	0.96	0.95	0.97	-0.94	-0.83
Job_Offers	0.24	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.96	1.00	0.97	0.99	-0.94	-0.82
Starting_Salary	0.20	0.95	0.95	0.96	0.94	0.94	0.97	0.95	0.95	0.97	1.00	0.97	-0.93	-0.85
Career_Satisfaction	0.24	0.96	0.97	0.97	0.96	0.96	0.96	0.97	0.97	0.99	0.97	1.00	-0.94	-0.81
Years_to_Promotion	-0.17	-0.96	-0.96	-0.96	-0.94	-0.93	-0.93	-0.94	-0.94	-0.94	-0.93	-0.94	1.00	0.83
Work_Life_Balance	-0.13	-0.80	-0.80	-0.83	-0.80	-0.79	-0.82	-0.83	-0.83	-0.82	-0.85	-0.81	0.83	1.00


DataPilot AI — AI Data Intellig
Ask Gemini
datapilot-ai.streamlit.app
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Upload Dataset
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Analysis Type:
Univariate
Univariate Analysis
Column:
Age
Chart:
Histogram
Age
High_School_GPA
SAT_Score
University_GPA
Internships_Completed
Projects_Completed
Certifications
Successfully loaded 400 rows × 19 columns from education_career_success.csv
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_wh...
Frequency
80
70
60
50
40
30


DataPilot AI — AI Data Intellig
Ask Gemini
datapilot-ai.streamlit.app
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Upload Dataset
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Analysis Type:
Univariate
Univariate Analysis
Column:
Age
Chart:
Histogram
Options
Bins:
30
Top N:
15
Histogram
Box Plot
Violin Plot
Bar Chart
Pie Chart
KDE Plot
Successfully loaded 400 rows × 19 columns from education_career_success.csv
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_wh...
Distribution of Age
Frequency
80
70
60
50
40
30


DataPilot AI — AI Data Intellig
Ask Gemini
datapilot-ai.streamlit.app
Analysis Type:
Univariate
Upload Dataset
educ
27.7K
Successfully loaded 400 rows × 19 columns from education_career_success.csv
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_wh...
Univariate Analysis
Column:
Age
Chart:
Histogram
Options
Distribution of Age
Frequency
Age
Download Chart
PNG
JPEG
SVG
HTML
Hi-Res


DataPilot AI — AI Data Intellig
Untitled document - Google Do
Ask Gemini
datapilot-ai.streamlit.app
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Upload Dataset
edu
27.7
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_w...
Force App Refresh
Made with by Nameer Sarookh DataPilot AI v2.0
Analysis Type:
Bivariate
Bivariate Analysis
X-axis:
Student_ID
Y-axis:
Age
Chart:
Scatter Plot
Options
Color by:
None
Add trendline
Age vs Student_ID
Age
Student_ID
Download Chart

DataPilot AI — AI Data Intellig | Streamlit Duplicate Element ID | Untitled document - Google D | Ask Gemini

datapilot-ai.streamlit.app

Upload Dataset

edu 27.7

File 400 × 19
Trust 79%
Missing 0
Numeric 14

Chart Theme

plotly_w...

Force App Refresh

Made with by Nameer Sarokh DataPilot AI v2.0

DataPilot AI

The AI Data Intelligence Copilot — Understand · Clean · Model · Decide

Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19

Overview | Visualize | AI Copilot | Statistics | Clean | ML Studio | Export

Analysis Type:

Time Series

Univariate
Bivariate
Multivariate
Time Series

No datetime columns auto-detected. Will attempt parsing.

Value column:

Age

Unknown datetime string format, unable to parse: S001

Chart:

Line with Range Slider

Options

DataPilot AI — AI Data Intellig | Streamlit Duplicate Element ID | Untitled document - Google D | Ask Gemini

datapilot-ai.streamlit.app

DataPilot AI

The AI Data Intelligence Copilot — Understand · Clean · Model · Decide

Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19

Overview | Visualize | AI Copilot | Statistics | Clean | ML Studio | Export

AI Data Copilot

AI powered by Groq LLM (llama-3.3-70b-versatile)

Quick Actions:

- What trends exist in this dataset?
- Which features matter most?
- Is this dataset reliable for ML?
- What cleaning steps do you recommend?
- Prepare ML-ready feature suggestions
- Are there any anomalies or outliers?

Ask anything about your data... (e.g. 'Why is revenue dropping?')

Dataset Context Summary

Shape: 400 rows × 19 columns

Numeric cols: Age, High_School_GPA, SAT_Score, University_GPA, Internships_Completed, Projects_Completed, Certifications, Soft_Skills_Score

Categorical cols: Student_ID, Gender, Field_of_Study, Current_Job_Level, Entrepreneurship

Missing values: 0

Upload Dataset

edu 27.7

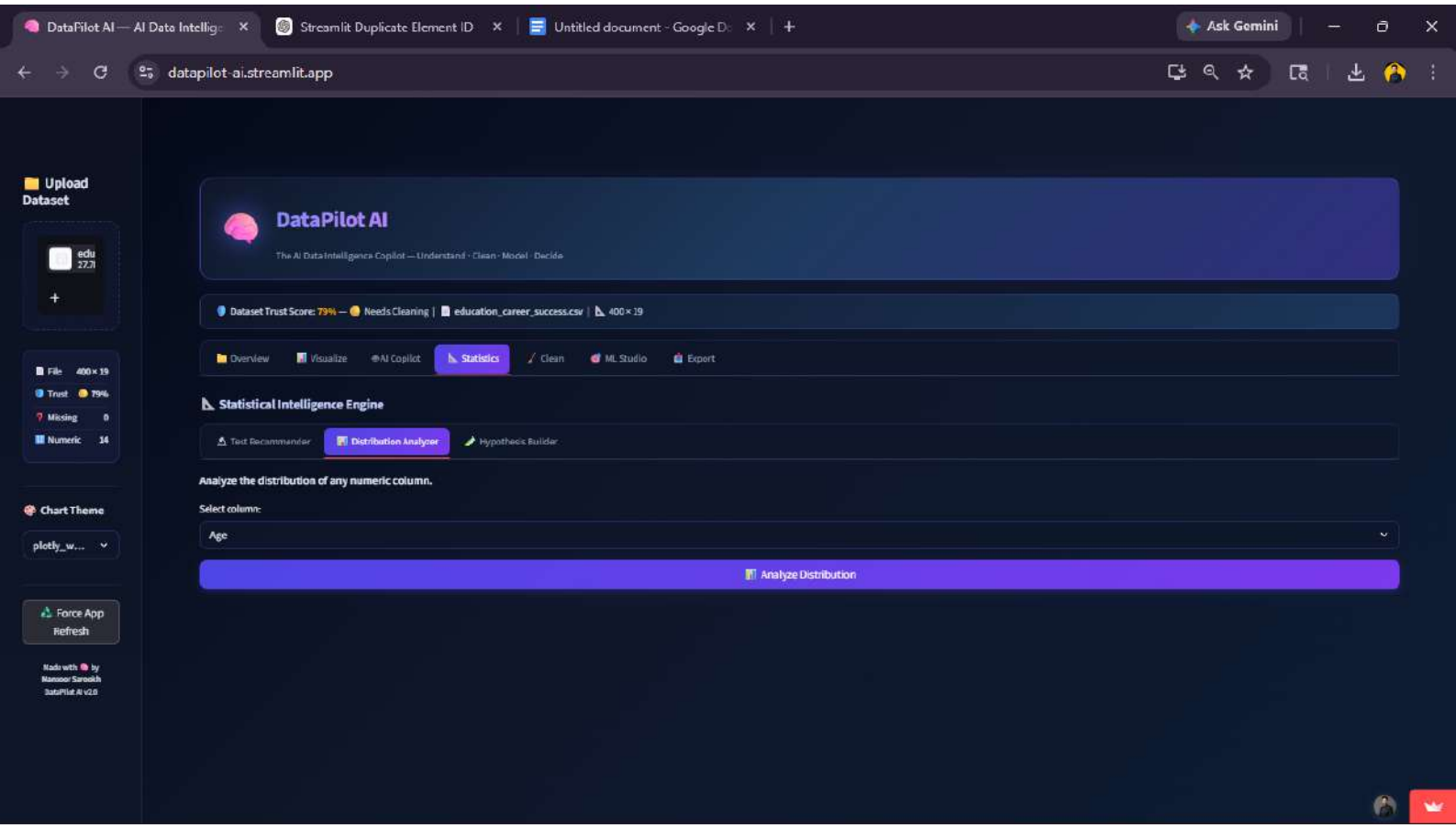
File 400 × 19
Trust 79%
Missing 0
Numeric 14

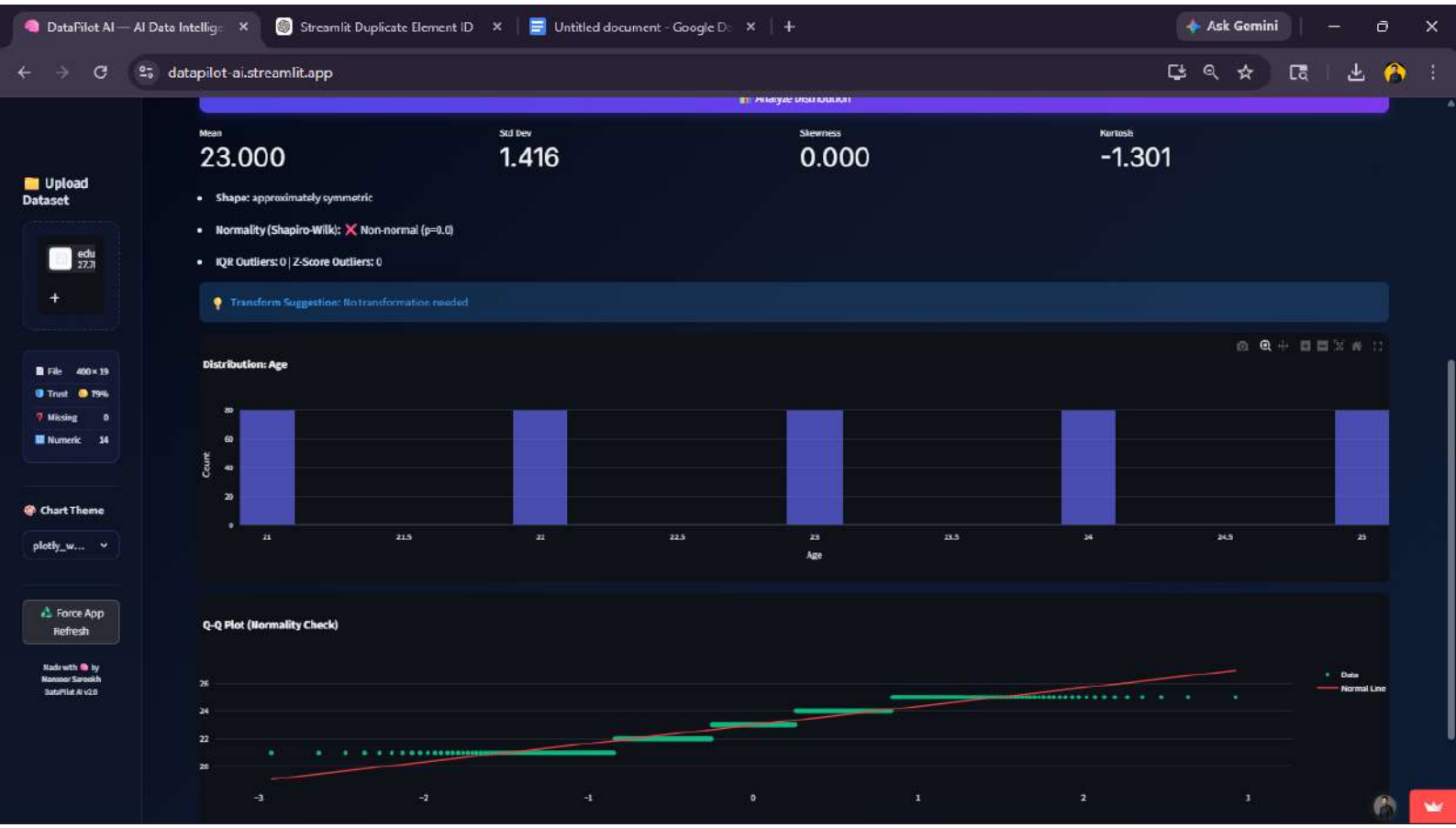
Chart Theme

plotly_w...

Force App Refresh

Made with by Nanoor Sarokh
DataPilot AI v2.0


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
edu
27.7
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_w...
Force App Refresh
Made with by Nanoor Sarokh
DataPilot AI v2.0
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Statistical Intelligence Engine
Test Recommender
Distribution Analyzer
Hypothesis Builder
Analyze the distribution of any numeric column.
Select column:
Age
Analyze Distribution


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Analyze Distribution
Mean
23.000
Std Dev
1.416
Skewness
0.000
Kurtosis
-1.301
Shape: approximately symmetric
Normality (Shapiro-Wilk): Non-normal (p=0.0)
IQR Outliers: 0 | Z-Score Outliers: 0
Transform Suggestion: No transformation needed
Distribution: Age
Count
Age
Q-Q Plot (Normality Check)
Data
Normal Line
Upload Dataset
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_w...
Force App Refresh

DataPilot AI — AI Data Intellig | Streamlit Duplicate Element ID | Untitled document - Google D | Ask Gemini

datapilot-ai.streamlit.app

Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19

Upload Dataset

edu 27.7

File 400 × 19
Trust 79%
Missing 0
Numeric 14

Chart Theme

plotly_w...

Force App Refresh

Made with by Nanoor Sarokh DataPilot AI v2.0

Overview | Visualize | AI Copilot | Statistics | Clean | ML Studio | Export

Statistical Intelligence Engine

Test Recommender | Distribution Analyzer | Hypothesis Builder

No-code hypothesis testing. Select variables and relationship — we do the rest.

Variable 1: Gender

Relationship: affects

Variable 2: Field_of_Study

Build & Test Hypothesis

H_0: 'Gender' has no effect on 'Field_of_Study' ($\beta = 0$)

H_1: 'Gender' significantly influences 'Field_of_Study' ($\beta \neq 0$)

Test: `Chi-Square Test of Independence`

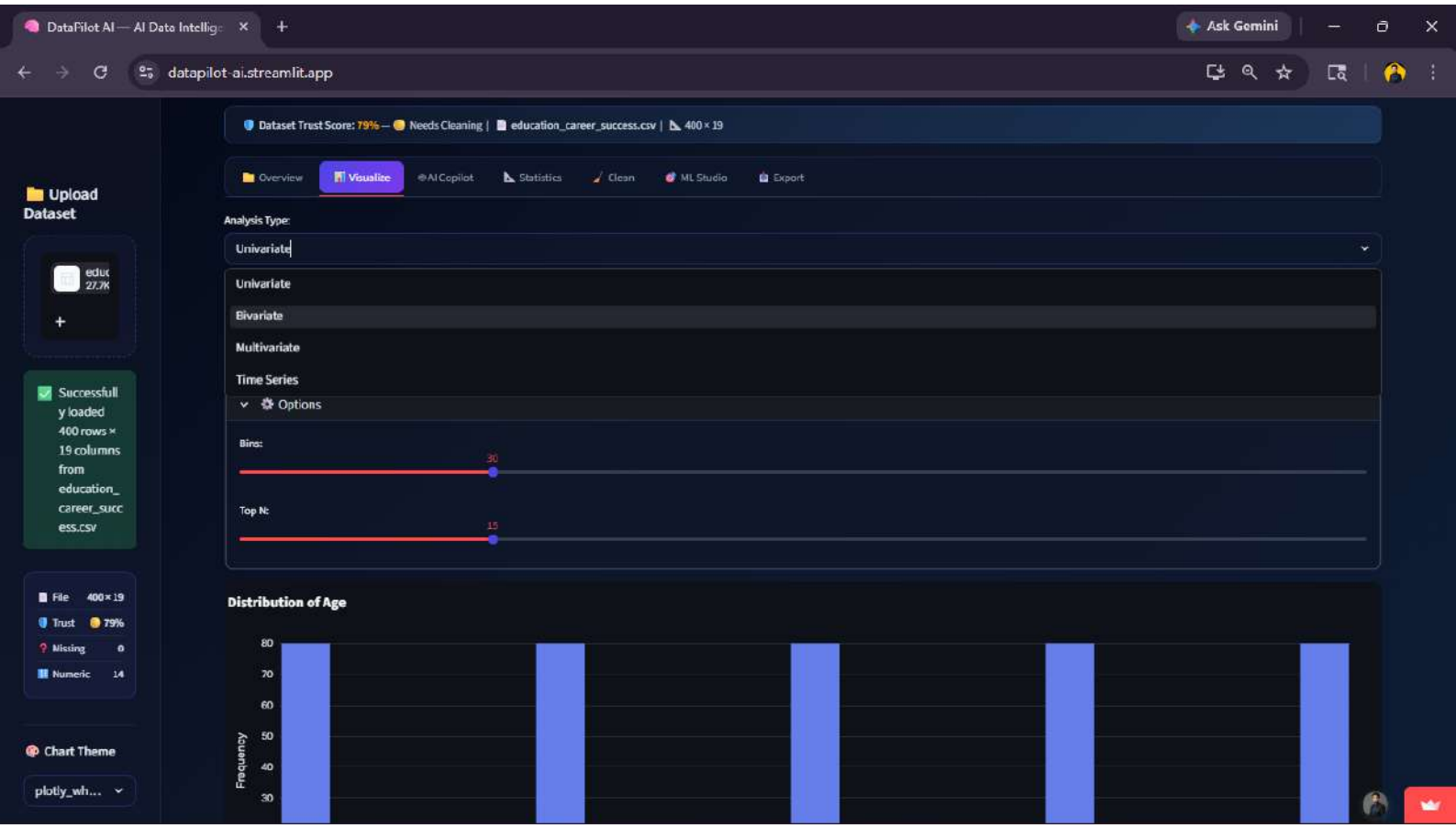
P-Value
N/A

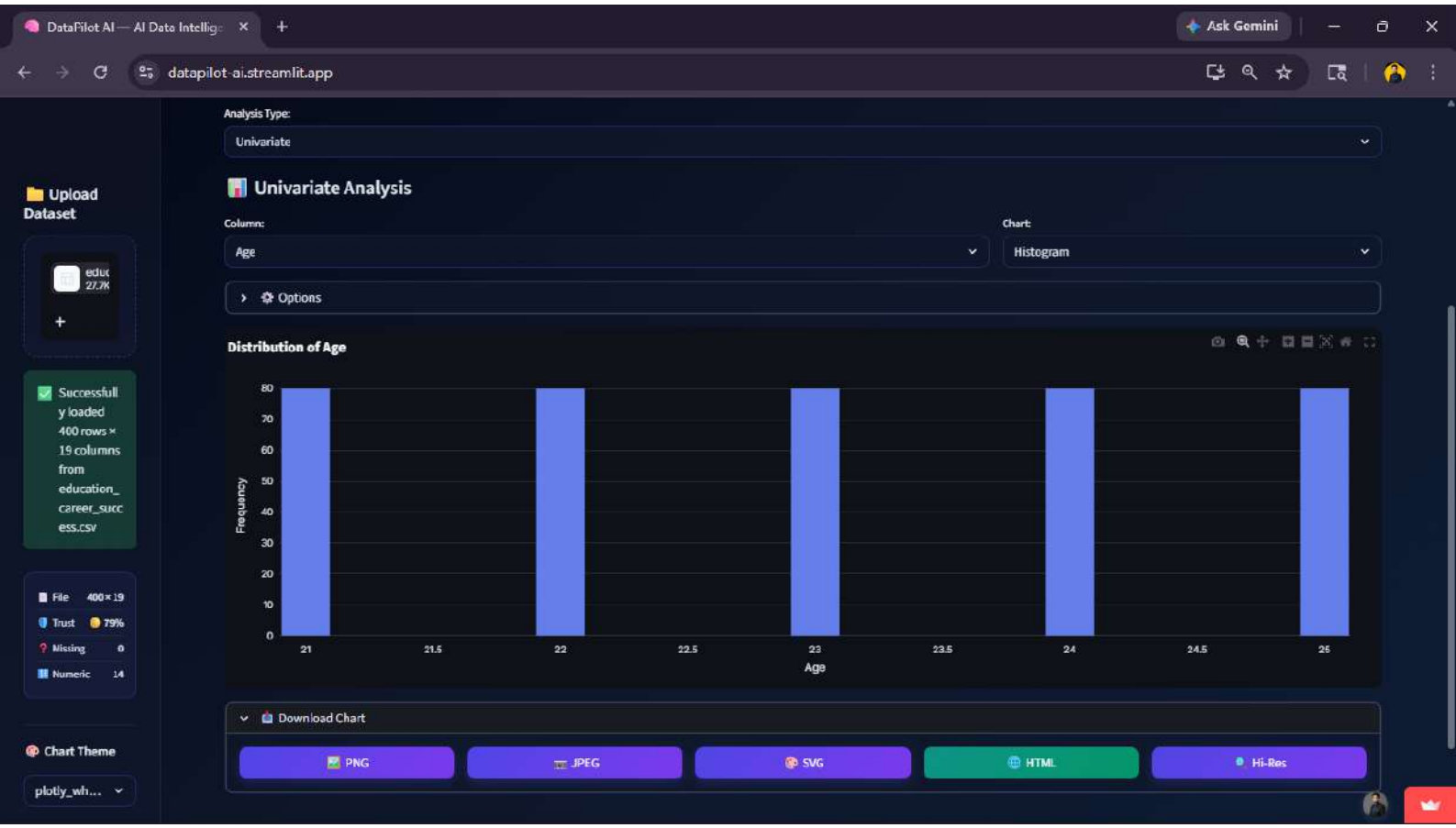
Decision
Reject H_0

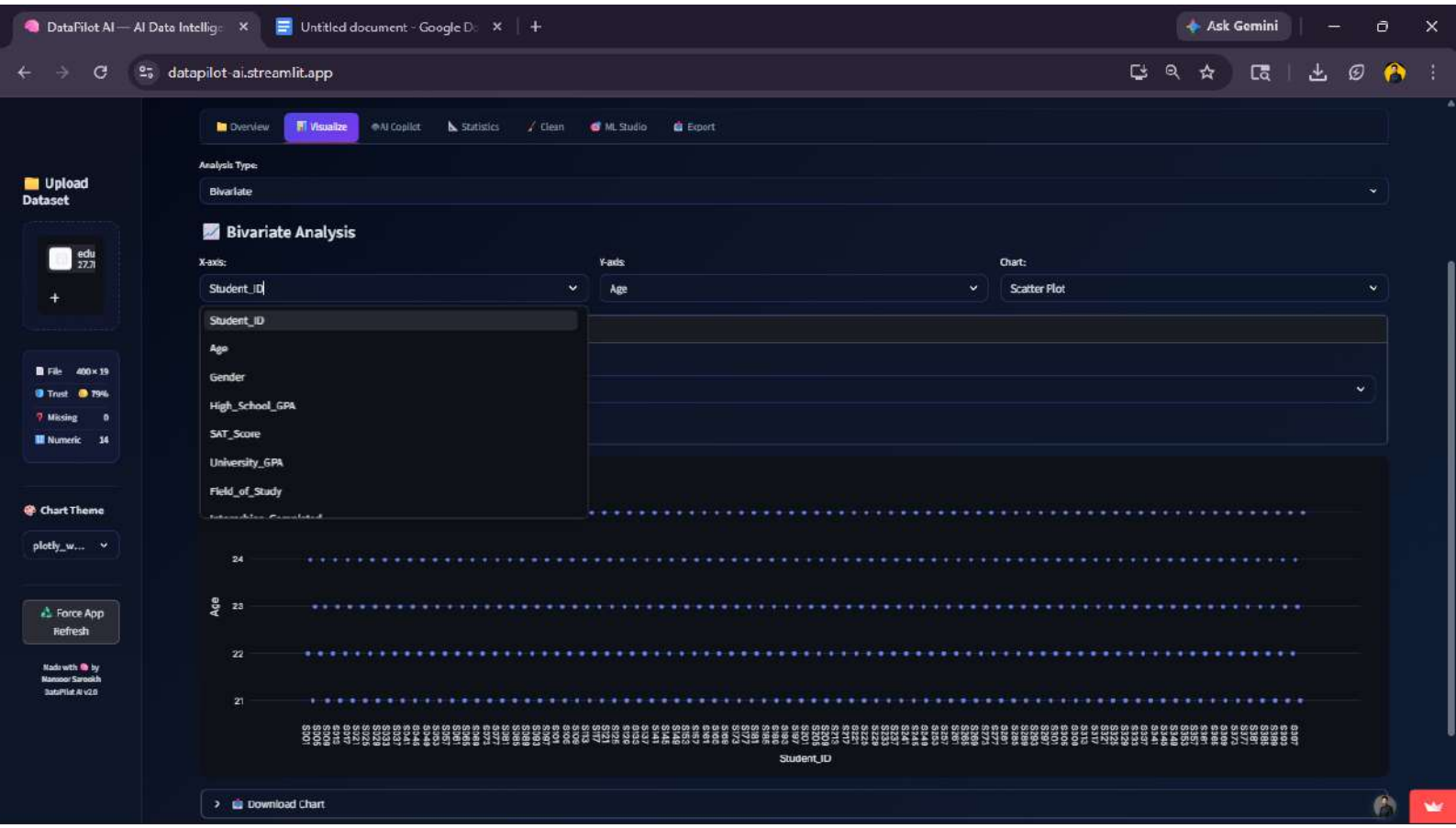
Reject H_0: 'Gender' significantly influences 'Field_of_Study' ($\beta \neq 0$) — statistically supported.

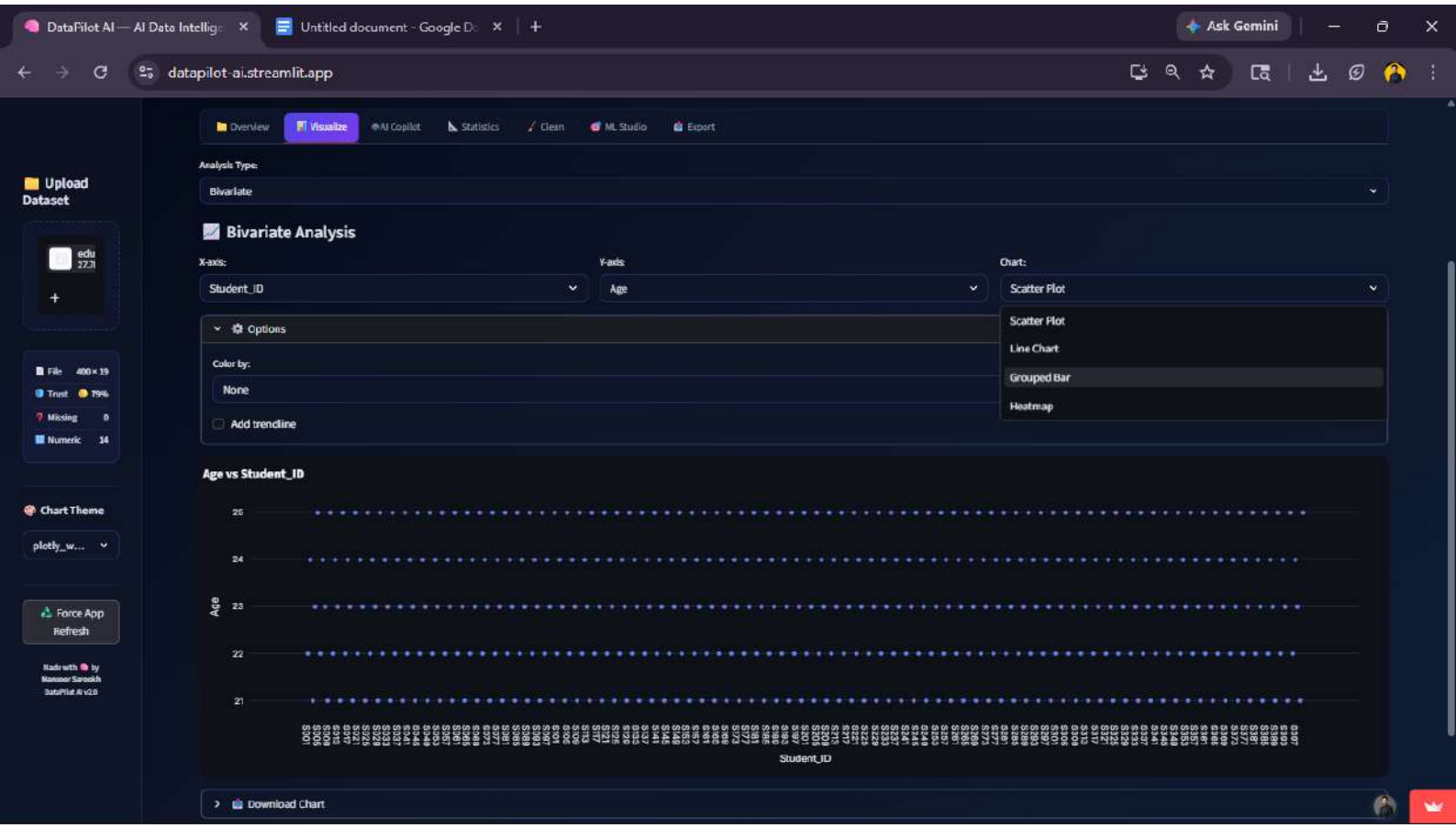
Significant association between 'Gender' and 'Field_of_Study' (χ^2=56.45, df=10, p=<0.001, V=0.376).

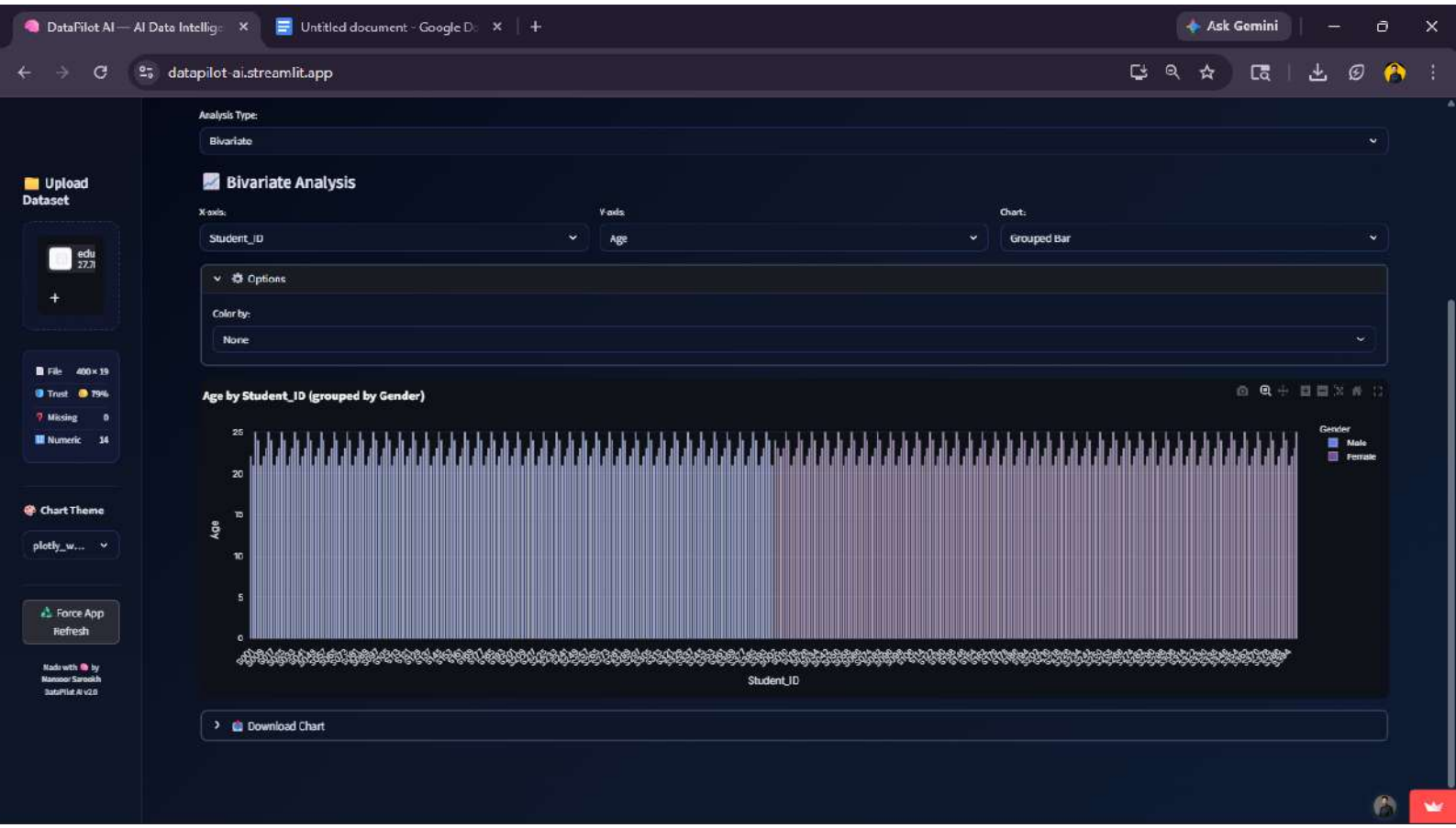
Next Step: Strong categorical association — consider interaction features for ML.

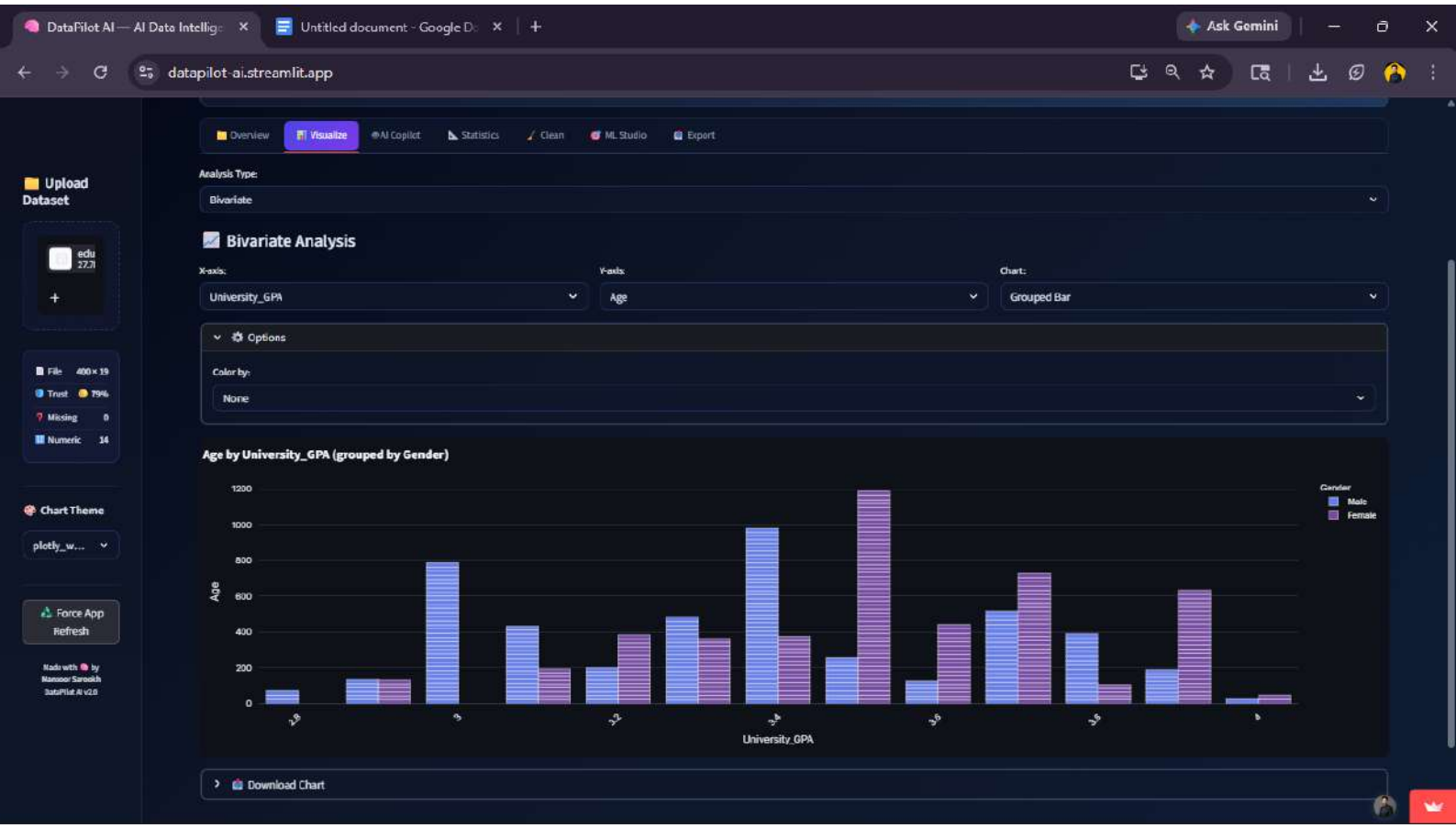

DataPilot AI — AI Data Intellig
Ask Gemini
datapilot-ai.streamlit.app
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Upload Dataset
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Analysis Type:
Univariate
Univariate
Bivariate
Multivariate
Time Series
Options
Bins:
30
Top N:
15
Successfully loaded 400 rows × 19 columns from education_career_success.csv
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_wh...
Distribution of Age
Frequency
80
70
60
50
40
30


DataPilot AI — AI Data Intellig
Ask Gemini
datapilot-ai.streamlit.app
Analysis Type:
Univariate
Upload Dataset
educ
27.7K
Successfully loaded 400 rows × 19 columns from education_career_success.csv
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_wh...
Univariate Analysis
Column:
Age
Chart:
Histogram
Options
Distribution of Age
Frequency
Age
Download Chart
PNG
JPEG
SVG
HTML
Hi-Res


DataPilot AI — AI Data Intellig
Untitled document - Google Do
Ask Gemini
datapilot-ai.streamlit.app
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Upload Dataset
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_w...
Force App Refresh
Made with by Nameer Sarookh DataPilot AI v2.0
Analysis Type:
Bivariate
Bivariate Analysis
X-axis:
Student_ID
Y-axis:
Age
Chart:
Scatter Plot
Student_ID
Age
Gender
High_School_GPA
SAT_Score
University_GPA
Field_of_Study
24
23
22
21
Age
Student_ID
Download Chart


DataPilot AI — AI Data Intellig
Untitled document - Google Do
Ask Gemini
datapilot-ai.streamlit.app
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Upload Dataset
edu
27.7
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_w...
Force App Refresh
Made with by Nameer Sarookh DataPilot AI v2.0
Analysis Type:
Bivariate
Bivariate Analysis
X-axis:
Student_ID
Y-axis:
Age
Chart:
Scatter Plot
Scatter Plot
Line Chart
Grouped Bar
Heatmap
Options
Color by:
None
Add trendline
Age vs Student_ID
Age
Student_ID
Download Chart


DataPilot AI — AI Data Intellige
Untitled document - Google Do
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_w...
Force App Refresh
Made with by Nameer Sarokih DataPilot AI v2.0
Analysis Type:
Bivariate
Bivariate Analysis
X-axis:
Student_ID
Y-axis:
Age
Chart:
Grouped Bar
Options
Color by:
None
Age by Student_ID (grouped by Gender)
Age
Student_ID
Gender
Male
Female
Download Chart


DataPilot AI — AI Data Intellig
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Upload Dataset
Analysis Type:
Bivariate
Bivariate Analysis
X-axis:
University_GPA
Y-axis:
Age
Chart:
Grouped Bar
Options
Color by:
None
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_w...
Force App Refresh
Age by University_GPA (grouped by Gender)
Gender
Male
Female
Age
University_GPA
Download Chart

DataPilot AI — AI Data Intellig | Streamlit Duplicate Element ID | Untitled document - Google D

Ask Gemini

datapilot-ai.streamlit.app

DataPilot AI

The AI Data Intelligence Copilot — Understand · Clean · Model · Decide

Upload Dataset

edu 27.7

File 400 × 19
Trust 79%
Missing 0
Numeric 14

Chart Theme

plotly_w...

Force App Refresh

Made with by Nanoor Sarokh
DataPilot AI v2.0

Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19

Overview | Visualize | AI Copilot | Statistics | Clean | ML Studio | Export

Statistical Intelligence Engine

Test Recommender | Distribution Analyzer | Hypothesis Builder

Auto-recommend the right statistical test for any two columns.

Variable 1: Student_ID

Variable 2: Age

Run Statistical Test

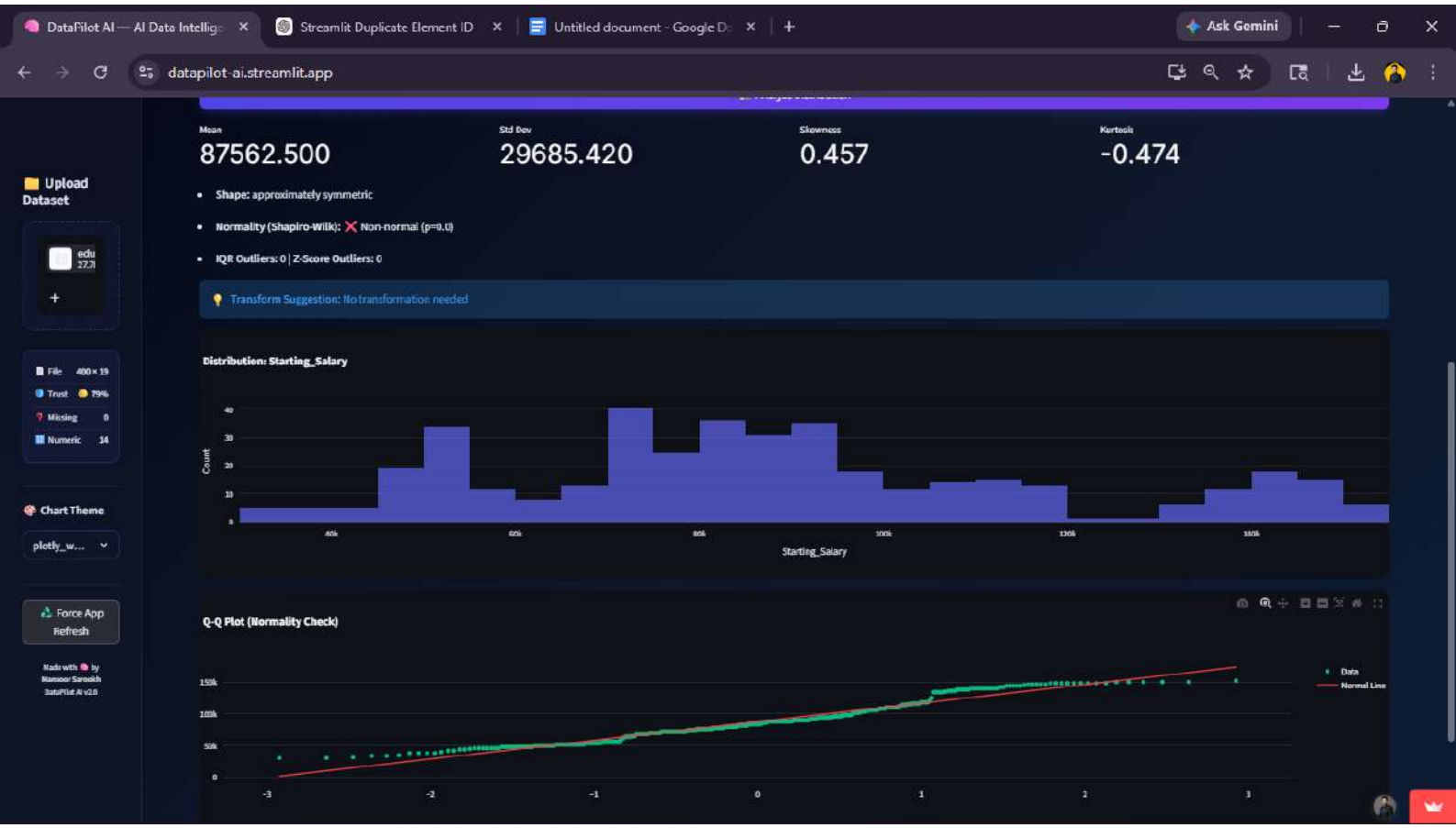
Test Selected: One-Way ANOVA

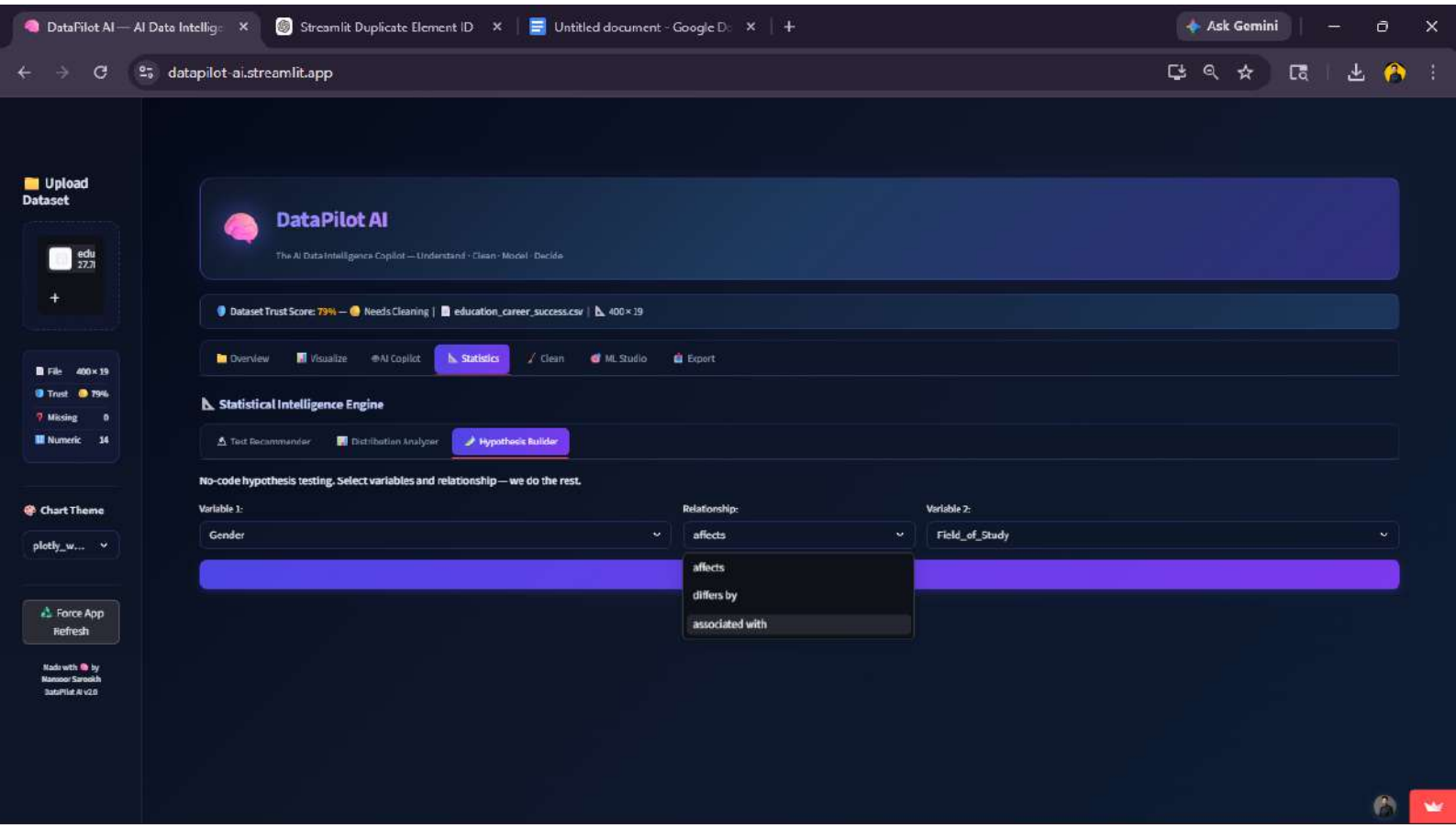
Test Statistic	P-Value	Significant?
nan	nan	❌ No

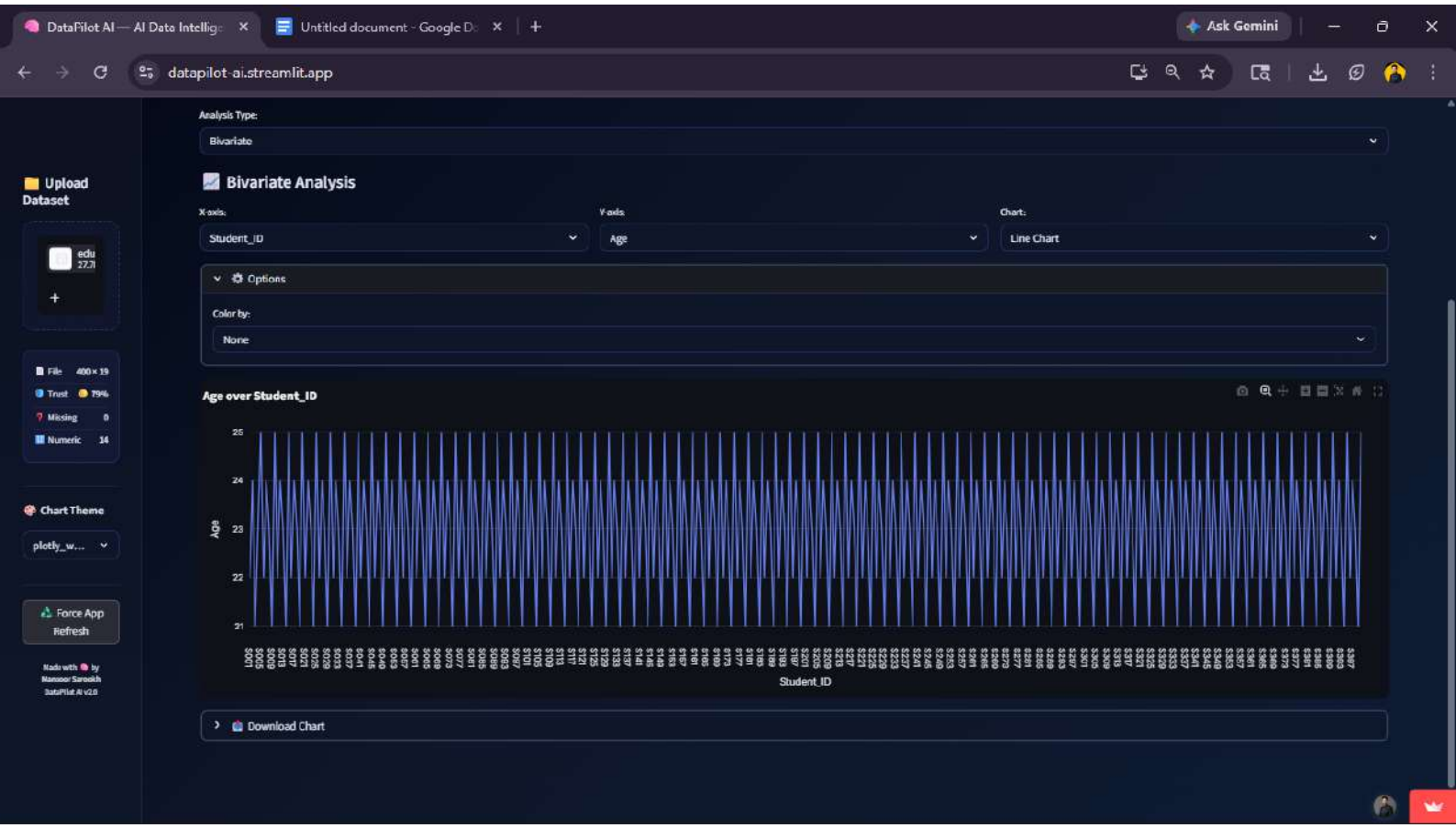
Assumptions Check

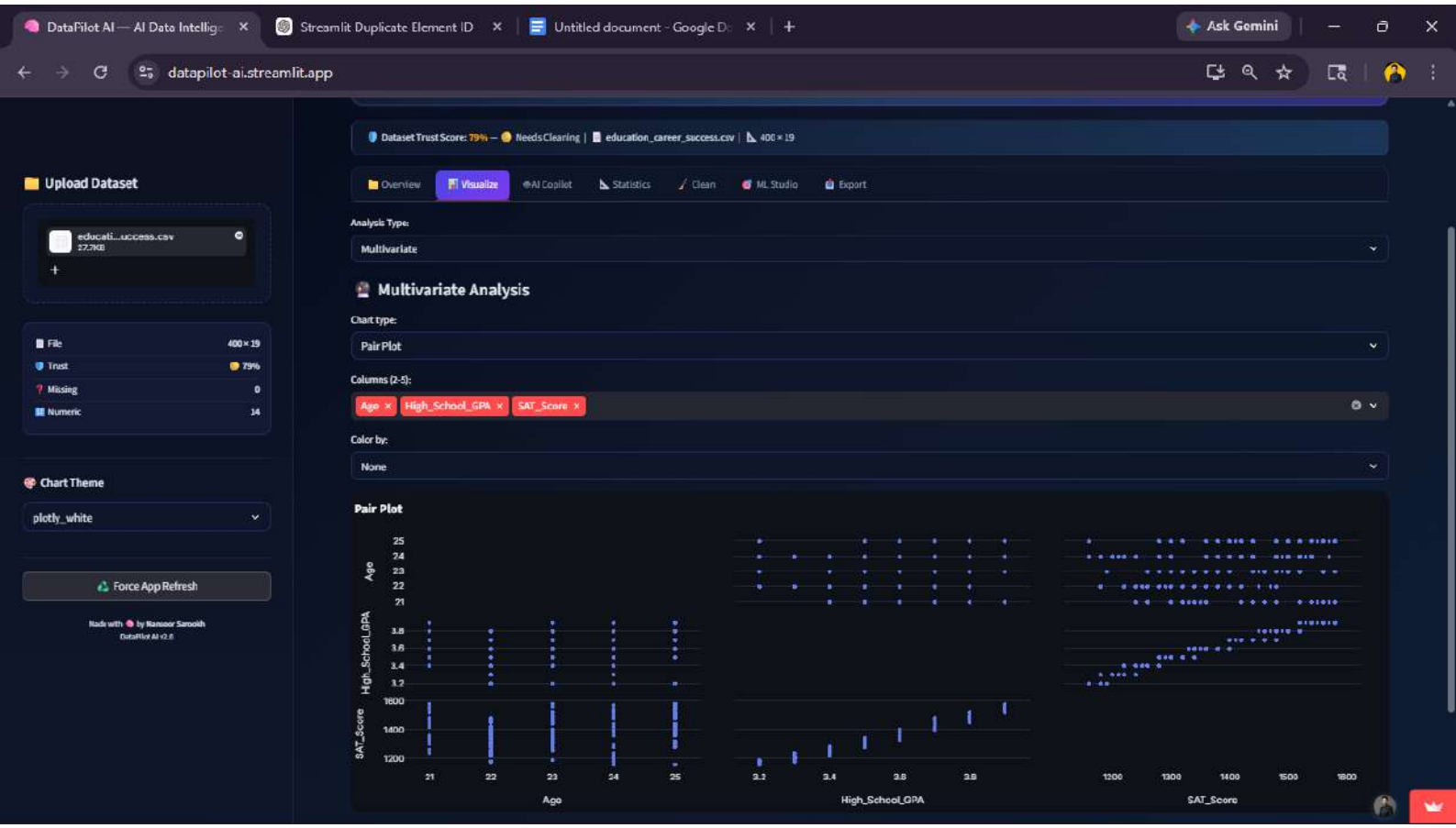
Interpretation: No significant difference in 'Age' across 'Student_ID' groups (One Way ANOVA: stat=nan, p=nan).

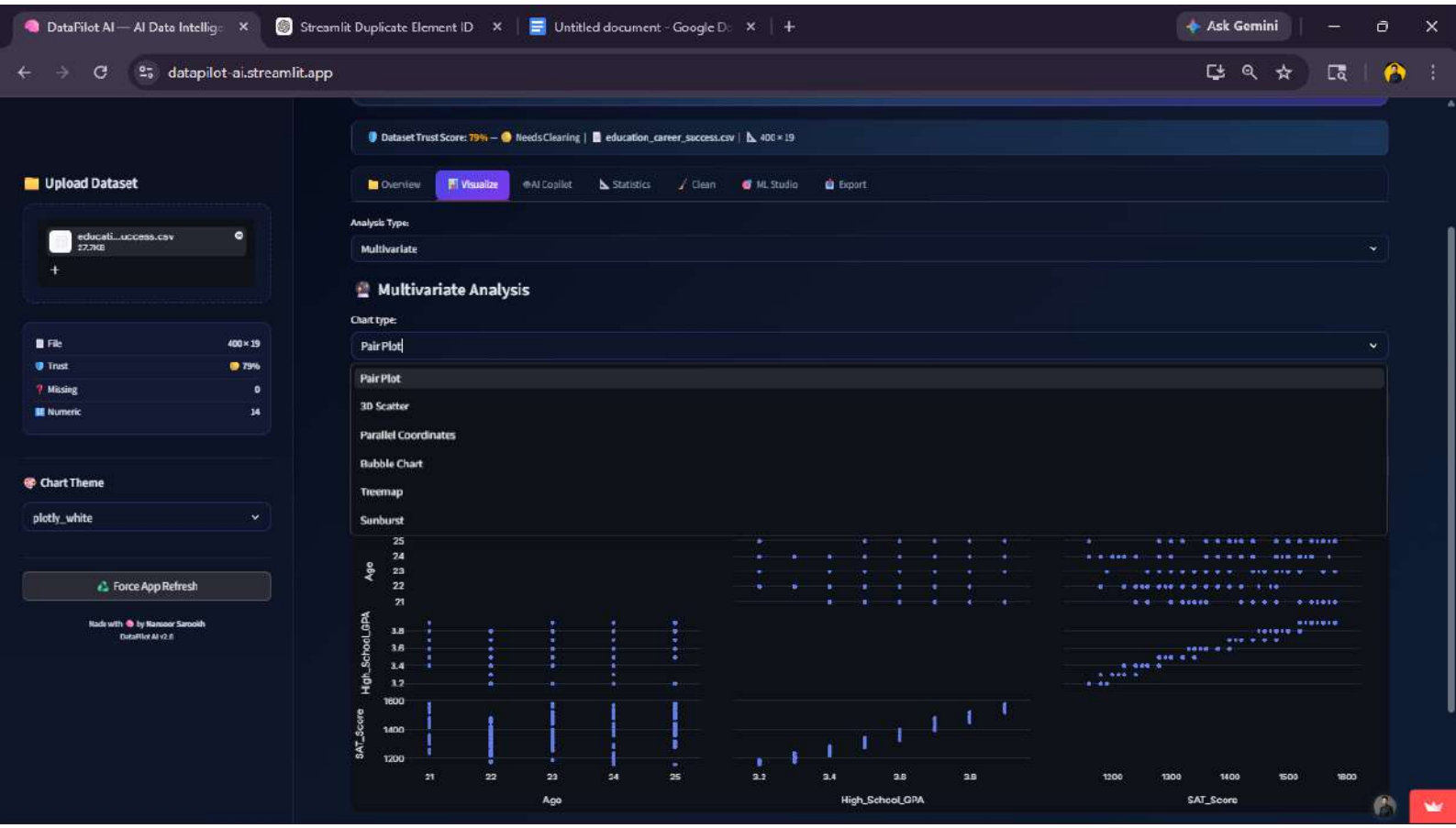
Recommendation: Grouping may not add predictive value.

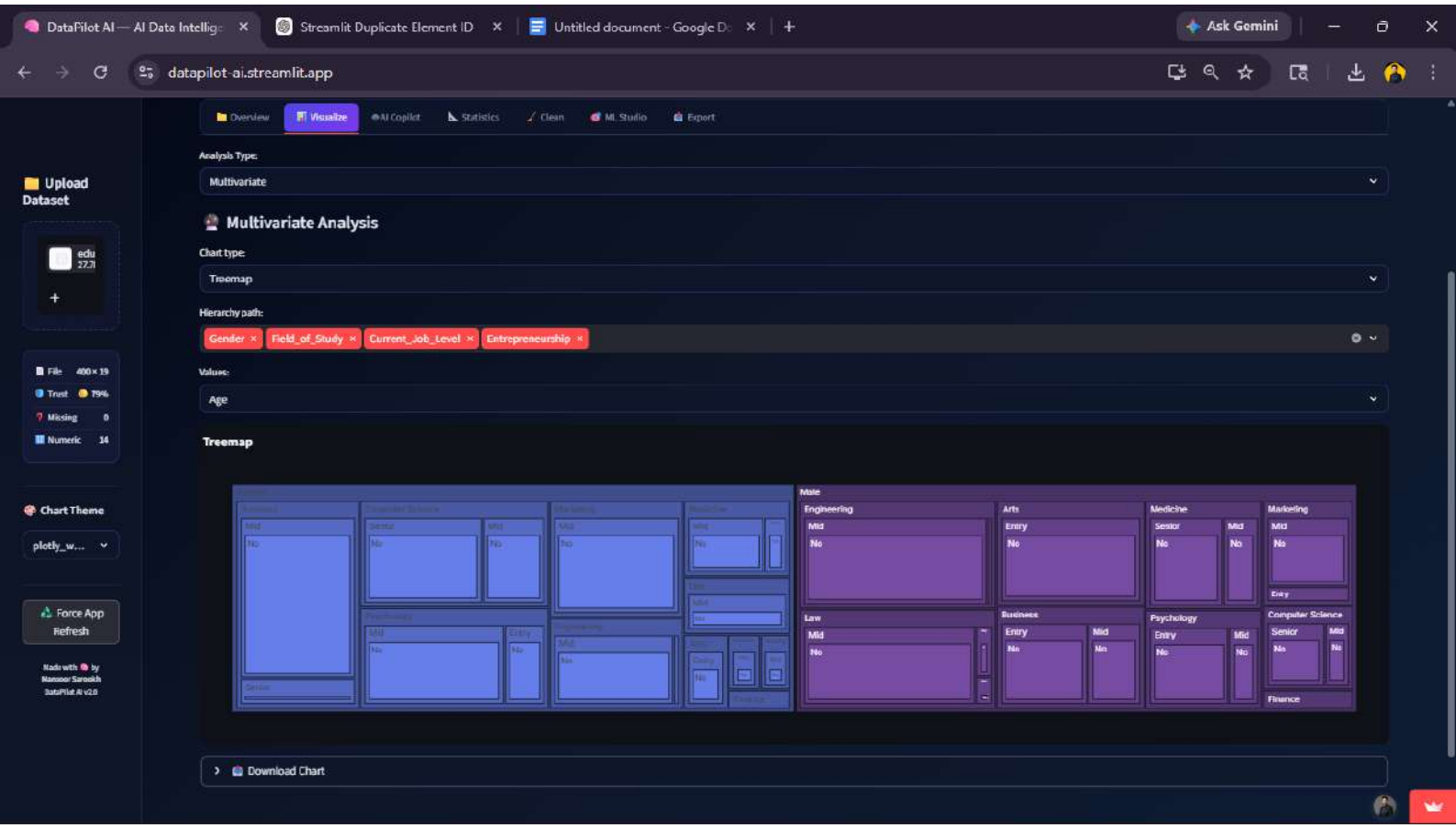

DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_w...
Force App Refresh
Made with by Nameer Sarokh DataPilot AI v2.0
Mean
87562.500
Std Dev
29685.420
Skewness
0.457
Kurtosis
-0.474
Shape: approximately symmetric
Normality (Shapiro-Wilk): Non-normal (p=0.0)
IQR Outliers: 0 | Z-Score Outliers: 0
Transform Suggestion: No transformation needed
Distribution: Starting_Salary
Count
Starting_Salary
Q-Q Plot (Normality Check)
Data
Normal Line


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Statistical Intelligence Engine
Test Recommender
Distribution Analyzer
Hypothesis Builder
No-code hypothesis testing. Select variables and relationship — we do the rest.
Variable 1:
Gender
Relationship:
affects
Variable 2:
Field_of_Study
affects
differs by
associated with
Chart Theme
plotly_w...
Force App Refresh
Made with by Nameer Sarookh
DataPilot AI v2.0


DataPilot AI — AI Data Intellig
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_w...
Force App Refresh
Made with by Nanoor Sarokh
DataPilot AI v2.0
Analysis Type:
Bivariate
Bivariate Analysis
X-axis:
Student_ID
Y-axis:
Age
Chart:
Line Chart
Options
Color by:
None
Age over Student_ID
Age
Student_ID
Download Chart


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Upload Dataset
educati...uccess.csv
27.7KB
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_white
Force App Refresh
Made with by Nansoor Sarookh
DataPilot AI v2.0
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Analysis Type:
Multivariate
Multivariate Analysis
Chart type:
Pair Plot
Columns (2-5):
Age
High_School_GPA
SAT_Score
Color by:
None
Pair Plot
Age
High_School_GPA
SAT_Score


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Upload Dataset
educati...uccess.csv
27.7KB
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_white
Force App Refresh
Made with by Nansoor Sarookh
DataPilot AI v2.0
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Analysis Type:
Multivariate
Multivariate Analysis
Chart type:
Pair Plot
Pair Plot
3D Scatter
Parallel Coordinates
Bubble Chart
Treemap
Sunburst
Age
High_School_GPA
SAT_Score


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Upload Dataset
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_w...
Force App Refresh
Analysis Type:
Multivariate
Multivariate Analysis
Chart type:
Treemap
Hierarchy path:
Gender
Field_of_Study
Current_Job_Level
Entrepreneurship
Values:
Age
Treemap
Male
Engineering
Arts
Medicine
Marketing
Law
Business
Psychology
Computer Science
Finance
Download Chart

DataPilot AI — AI Data Intellig | Streamlit Duplicate Element ID | Untitled document - Google D | Ask Gemini

datapilot-ai.streamlit.app

1. The mean number of internships completed is 2.5, and the mean number of projects completed is 6.235, indicating that students have more project experience than internship experience.
2. The mean soft skills score (7.933) is higher than the mean networking score (6.933), suggesting that students may have stronger soft skills than networking skills.
3. The mean starting salary is not provided, but the sample data shows a range of starting salaries from 40,000 to 120,000, indicating a potential relationship between university GPA, internships, and starting salary.

⚠️ **Risk:** The dataset may not capture the full range of experiences and outcomes for all students, potentially leading to biased trends.

✅ **Action:** Apply correlation analysis and visualization techniques (e.g., scatter plots, heatmaps) to explore relationships between variables, such as university GPA, internships, and starting salary, to identify more trends and patterns in the data.

Prepare ML-ready feature suggestions

📊 **Insight:**

1. The dataset contains a mix of numeric and categorical features, with 14 numeric columns and 5 categorical columns.
2. The numeric columns have varying means and standard deviations, such as "Age" (mean: 23.0, std: 1.416) and "University_GPA" (mean: 3.441, std: 0.286).
3. The categorical columns, such as "Field_of_Study" and "Current_Job_Level", can be encoded using techniques like one-hot encoding or label encoding.

⚠️ **Risk:** The dataset may require handling of categorical variables and potential scaling/normalization of numeric variables to ensure ML model performance.

✅ **Action:**

- Create new features, such as:
  - "GPA_Ratio" (University_GPA / High_School_GPA)
  - "Experience_Score" (Internships_Completed + Projects_Completed)
- Encode categorical variables using one-hot encoding or label encoding
- Scale/normalize numeric variables using techniques like Standard Scaler or Min-Max Scaler
- Consider feature selection techniques to reduce dimensionality and improve model performance

Ask anything about your data... (e.g. 'Why is revenue dropping?')

Clear Chat History

Dataset Context Summary

Upload Dataset

edu 27.7

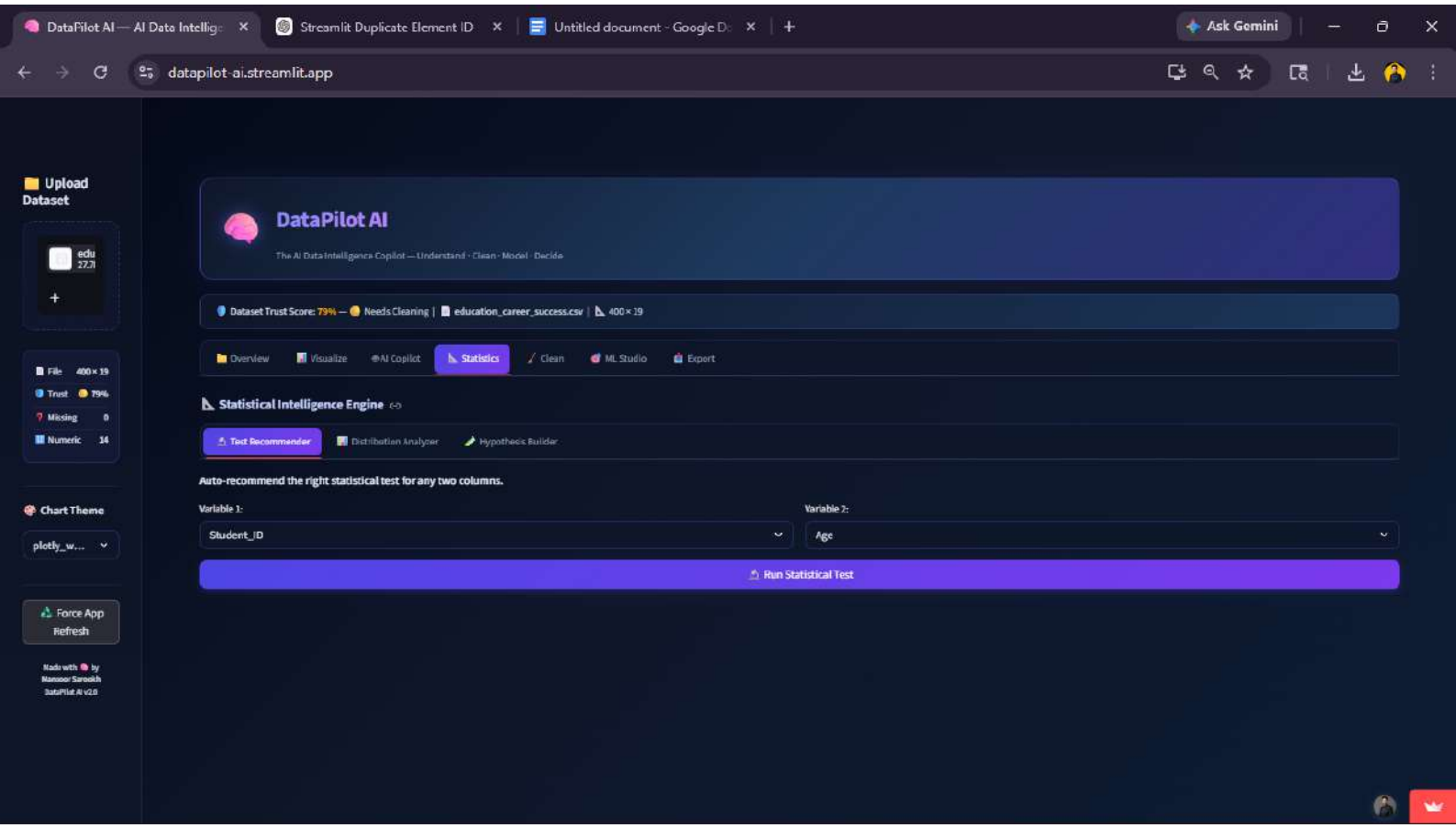
File 400 × 19
Trust 79%
Missing 0
Numeric 14

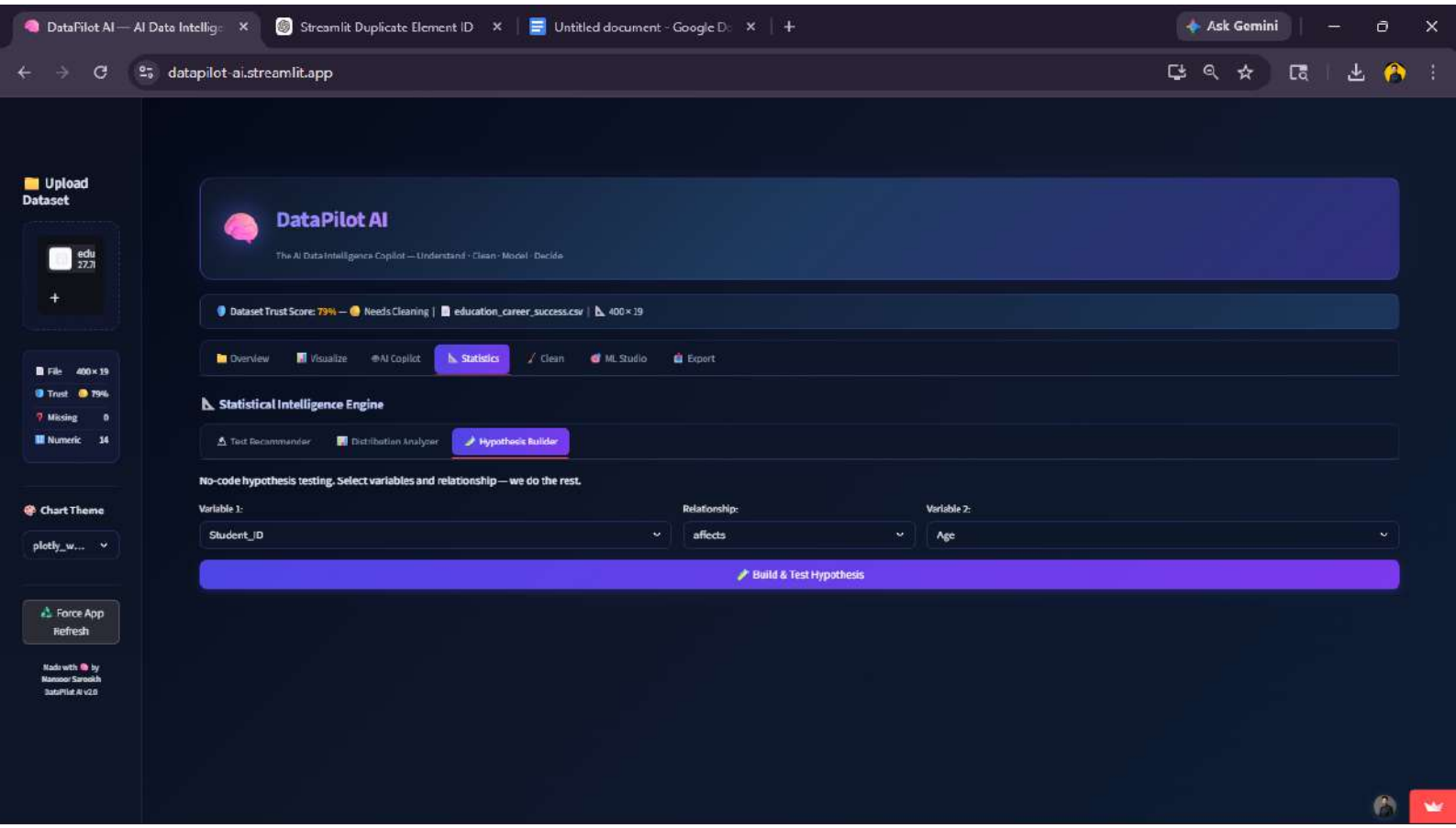
Chart Theme

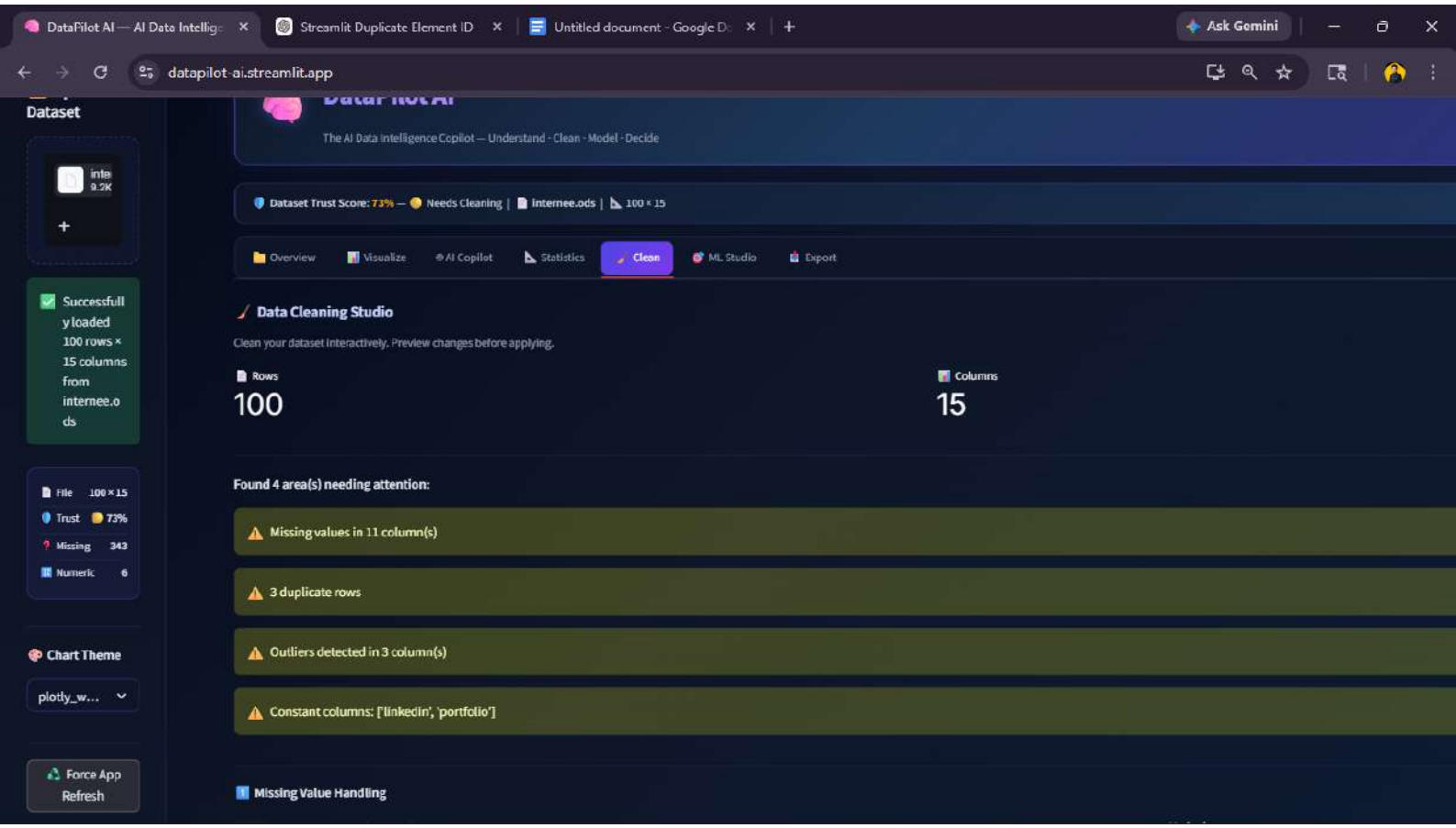
plotly_w...

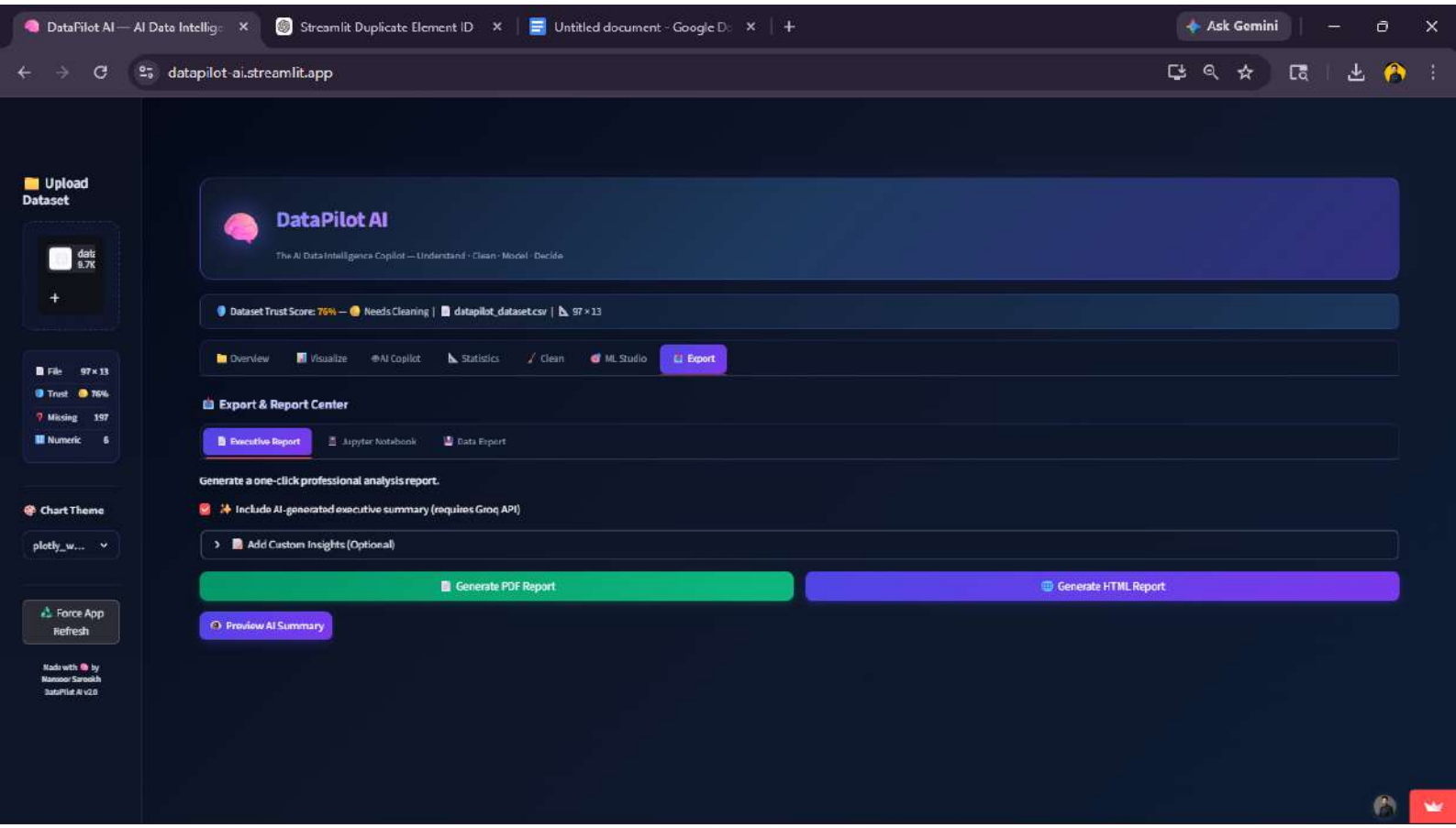
Force App Refresh

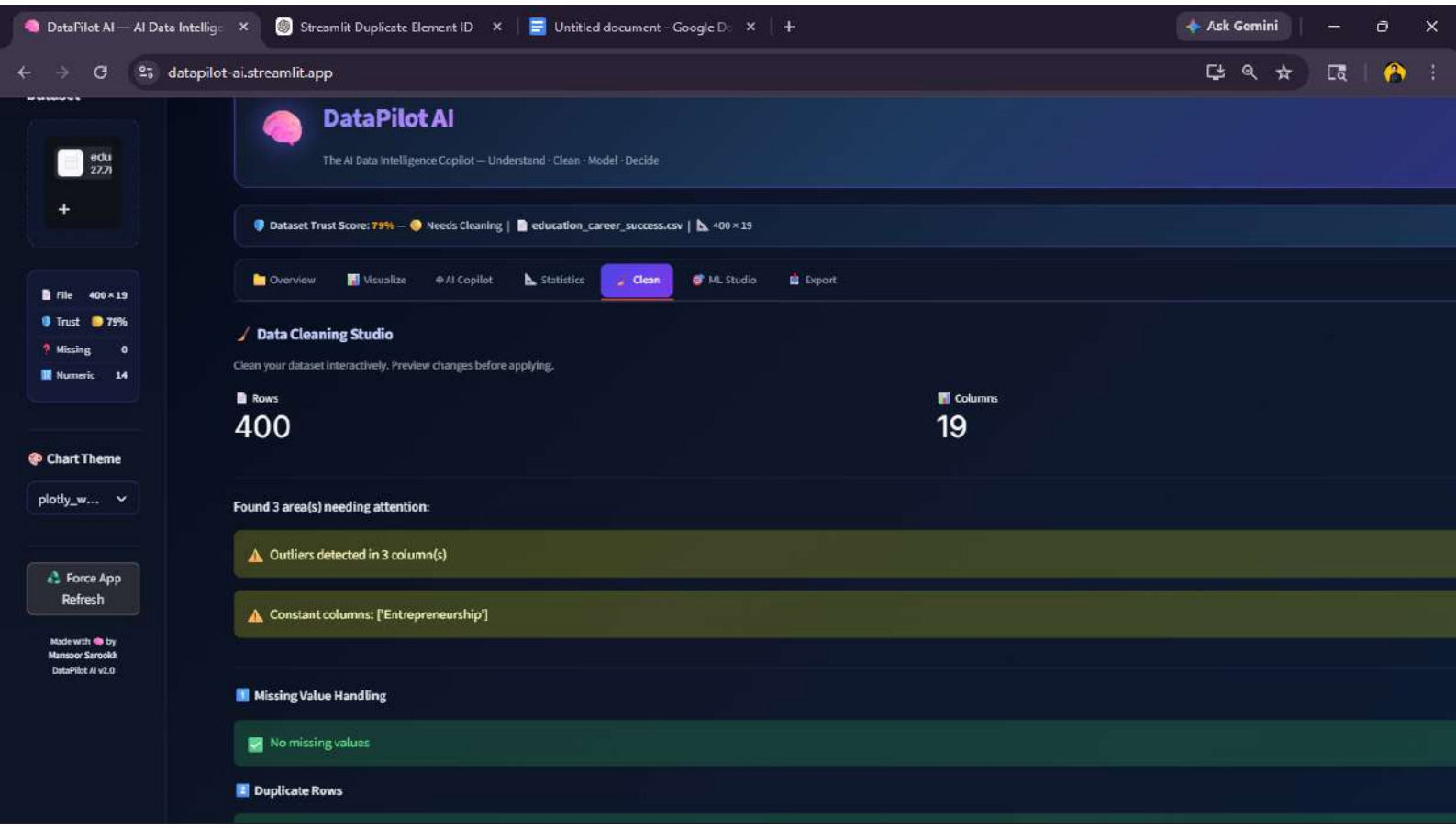
Made with by Nanoor Sarokh
DataPilot AI v2.0


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
edu
27.7
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_w...
Force App Refresh
Made with by Nanoor Sarokh
DataPilot AI v2.0
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Statistical Intelligence Engine
Test Recommender
Distribution Analyzer
Hypothesis Builder
Auto-recommend the right statistical test for any two columns.
Variable 1:
Student_ID
Variable 2:
Age
Run Statistical Test


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
edu
27.7
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_w...
Force App Refresh
Made with by Nanoor Sarokh
DataPilot AI v2.0
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Statistical Intelligence Engine
Test Recommender
Distribution Analyzer
Hypothesis Builder
No-code hypothesis testing. Select variables and relationship — we do the rest.
Variable 1:
Student_ID
Relationship:
affects
Variable 2:
Age
Build & Test Hypothesis


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Dataset
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 73% — Needs Cleaning | internee.ods | 100 × 15
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Successfully loaded 100 rows × 15 columns from internee.ods
Data Cleaning Studio
Clean your dataset interactively. Preview changes before applying.
Rows
100
Columns
15
File 100 × 15
Trust 73%
Missing 343
Numeric 6
Found 4 area(s) needing attention:
Missing values in 11 column(s)
3 duplicate rows
Outliers detected in 3 column(s)
Constant columns: ['linkedin', 'portfolio']
Chart Theme
plotly_w...
Force App Refresh
Missing Value Handling


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
data 8.7K
File 97 × 13
Trust 76%
Missing 197
Numeric 6
Chart Theme
plotly_w...
Force App Refresh
Made with by Nanoor Sarokh
DataPilot AI v2.0
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 76% — Needs Cleaning | datapilot_dataset.csv | 97 × 13
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Export & Report Center
Executive Report
Jupyter Notebook
Data Export
Generate a one-click professional analysis report.
Include AI-generated executive summary (requires Groq API)
Add Custom Insights (Optional)
Generate PDF Report
Generate HTML Report
Preview AI Summary


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 79% — Needs Cleaning | education_career_success.csv | 400 × 19
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
File 400 × 19
Trust 79%
Missing 0
Numeric 14
Chart Theme
plotly_w...
Force App Refresh
Made with by Mansoor Sarookh DataPilot AI v2.0
Data Cleaning Studio
Clean your dataset interactively. Preview changes before applying.
Rows
400
Columns
19
Found 3 area(s) needing attention:
Outliers detected in 3 column(s)
Constant columns: ['Entrepreneurship']
Missing Value Handling
No missing values
Duplicate Rows

DataPilot AI — AI Data Intellig | Streamlit Duplicate Element ID | Untitled document - Google D

Ask Gemini

datapilot-ai.streamlit.app

Dataset

edu 277

File 400 × 19

Trust 79%

Missing 0

Numeric 14

Chart Theme

plotly_w...

Force App Refresh

Made with by Mansoor Sarookh DataPilot AI v2.0

No missing values

2 Duplicate Rows

No duplicates

3 Constant Columns

Drop constant columns: ['Entrepreneurship']

4 Outlier Handling

Outliers detected in: ['Internships_Completed', 'Certifications', 'Work_Life_Balance']

Outlier strategy:

iqr_clip

5 Categorical Encoding

Encoding method:

None

6 Feature Scaling

Scaling method:

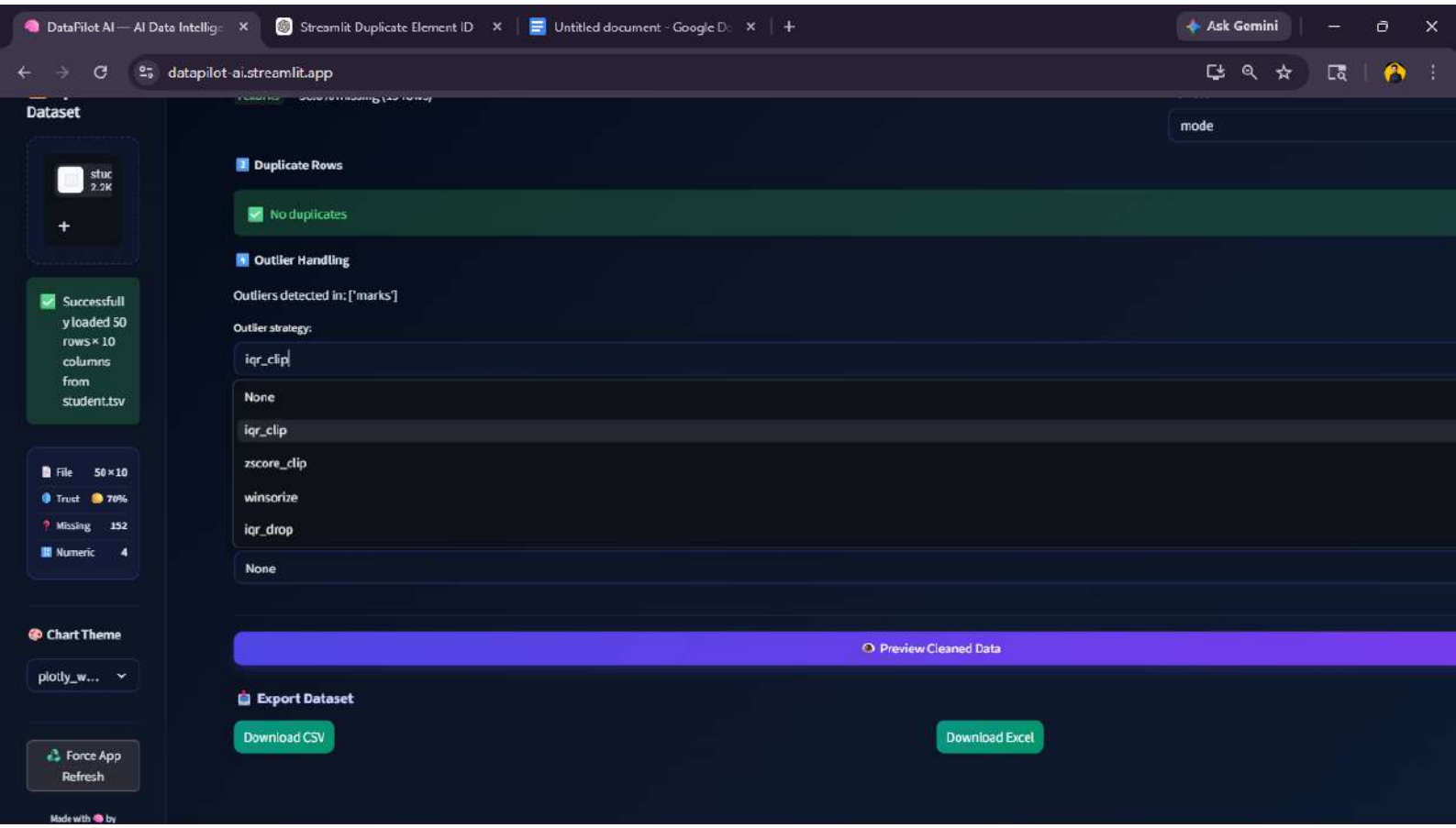
None

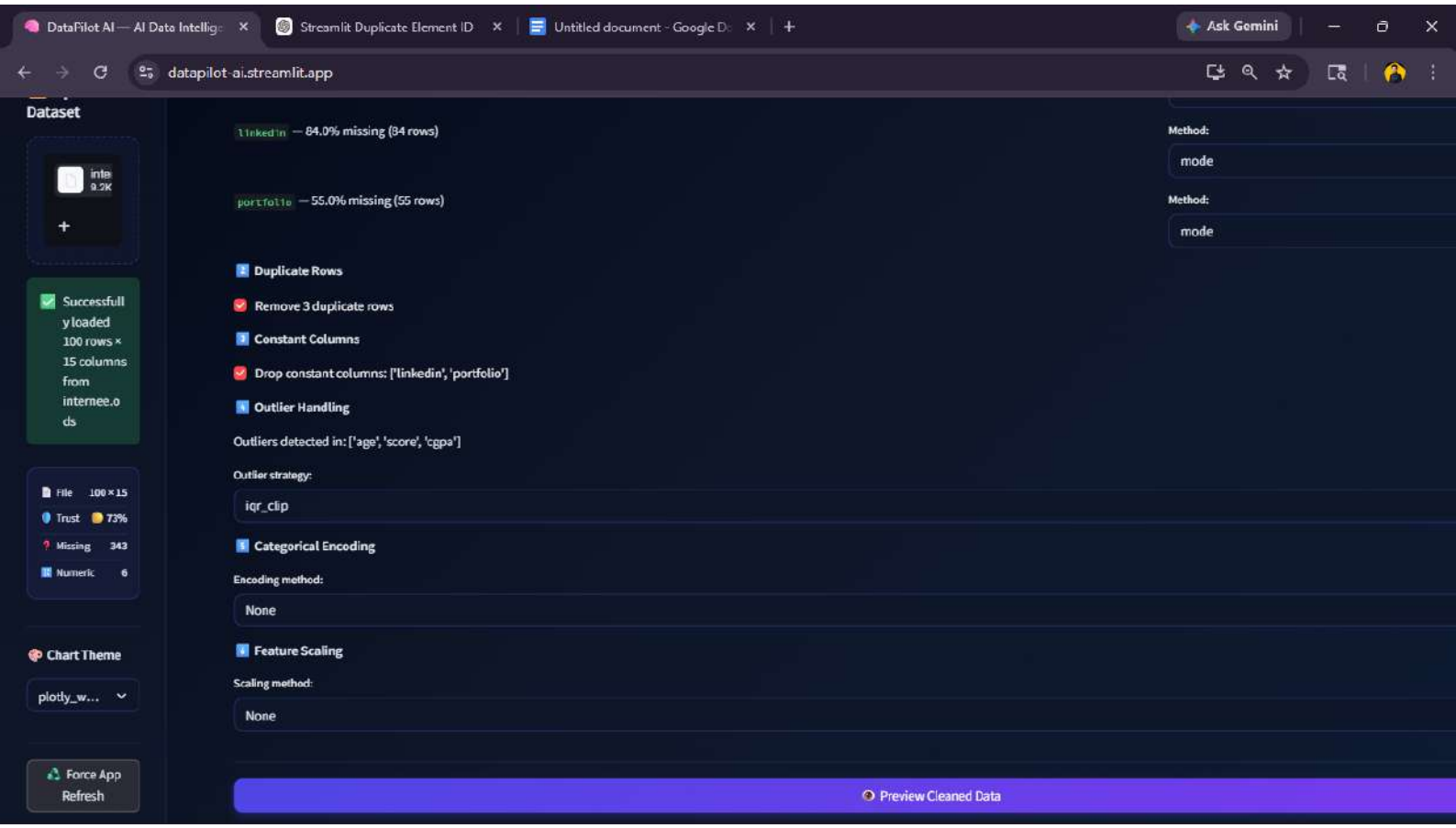
Preview Cleaned Data

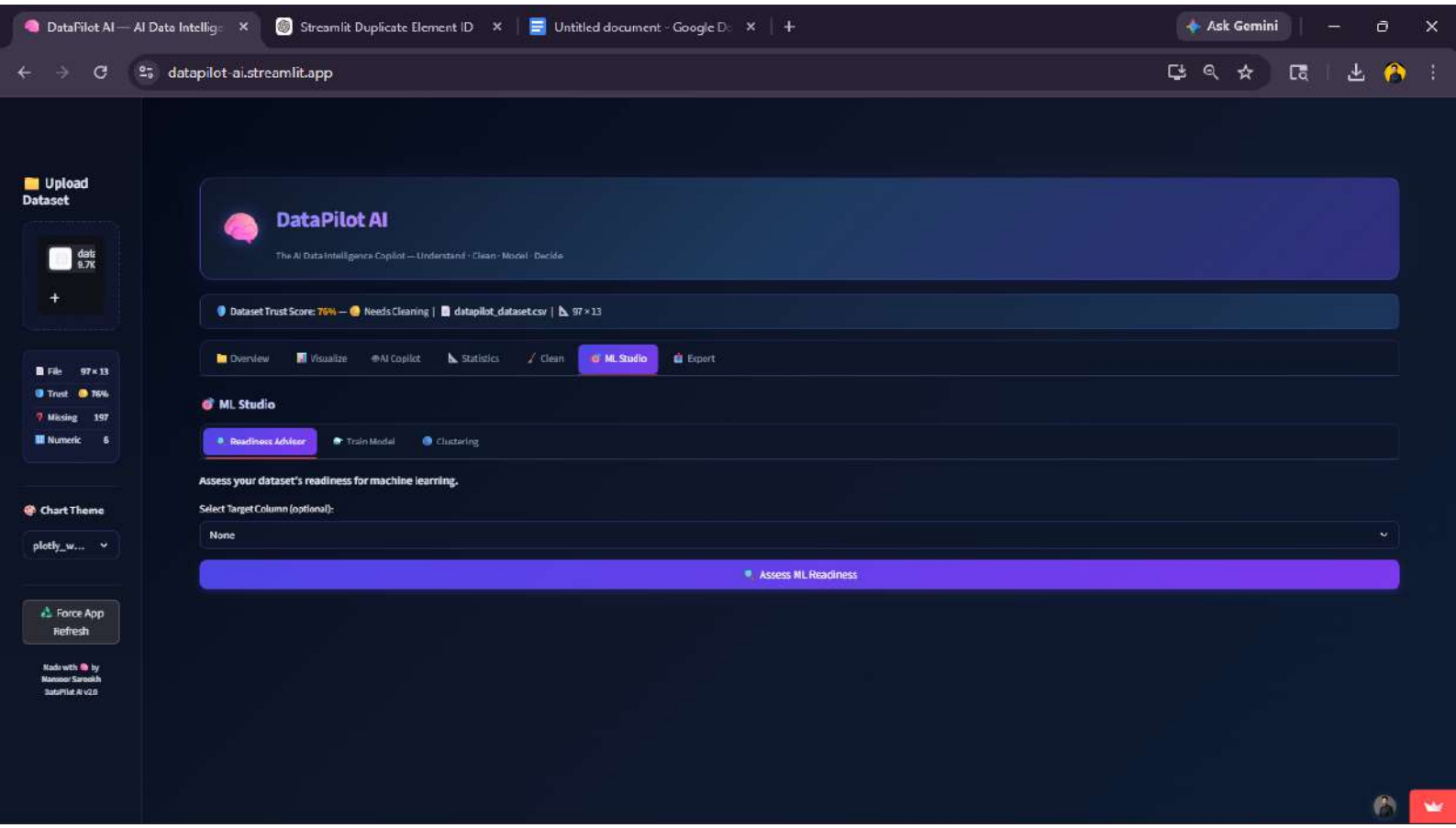
Export Dataset

Download CSV

Download Excel


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Dataset
stuc
2.2K
Successfully loaded 50 rows × 10 columns from student.tsv
File 50 × 10
Trust 70%
Missing 152
Numeric 4
Chart Theme
plotly_w...
Force App Refresh
Made with by
mode
Duplicate Rows
No duplicates
Outlier Handling
Outliers detected in: ['marks']
Outlier strategy:
iqr_clip
None
iqr_clip
zscore_clip
winsorize
iqr_drop
None
Preview Cleaned Data
Export Dataset
Download CSV
Download Excel


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Dataset
Successfully loaded 100 rows × 15 columns from internee.ods
File 100 × 15
Trust 73%
Missing 343
Numeric 6
Chart Theme
plotly_w...
Force App Refresh
linkedin — 84.0% missing (84 rows)
portfolio — 55.0% missing (55 rows)
Method:
mode
Method:
mode
2 Duplicate Rows
Remove 3 duplicate rows
3 Constant Columns
Drop constant columns: ['linkedin', 'portfolio']
4 Outlier Handling
Outliers detected in: ['age', 'score', 'cgpa']
Outlier strategy:
iqr_clip
5 Categorical Encoding
Encoding method:
None
6 Feature Scaling
Scaling method:
None
Preview Cleaned Data


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
data 8.7K
File 97 × 13
Trust 76%
Missing 197
Numeric 6
Chart Theme
plotly_w...
Force App Refresh
Made with by Nanoor Sarokh DataPilot AI v2.0
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 76% — Needs Cleaning | datapilot_dataset.csv | 97 × 13
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
ML Studio
Readiness Advisor
Train Model
Clustering
Assess your dataset's readiness for machine learning.
Select Target Column (optional):
None
Assess ML Readiness

DataPilot AI — AI Data Intellig | Streamlit Duplicate Element ID | Untitled document - Google D | Ask Gemini

datapilot-ai.streamlit.app

DataPilot AI

The AI Data Intelligence Copilot — Understand · Clean · Model · Decide

Upload Dataset

data 8.7K

+

File 97 × 13
Trust 76%
Missing 197
Numeric 6

Chart Theme

plotly_w...

Force App Refresh

Made with by Nanoor Sarokh DataPilot AI v2.0

Dataset Trust Score: 76% — Needs Cleaning | datapilot_dataset.csv | 97 × 13

Overview | Visualize | AI Copilot | Statistics | Clean | ML Studio | Export

ML Studio

Readiness Adviser | Train Model | Clustering

Train an ML model directly in your browser.

Target Column:

cgpa

Detected Problem Type: Classification (6 unique values)

Algorithm:

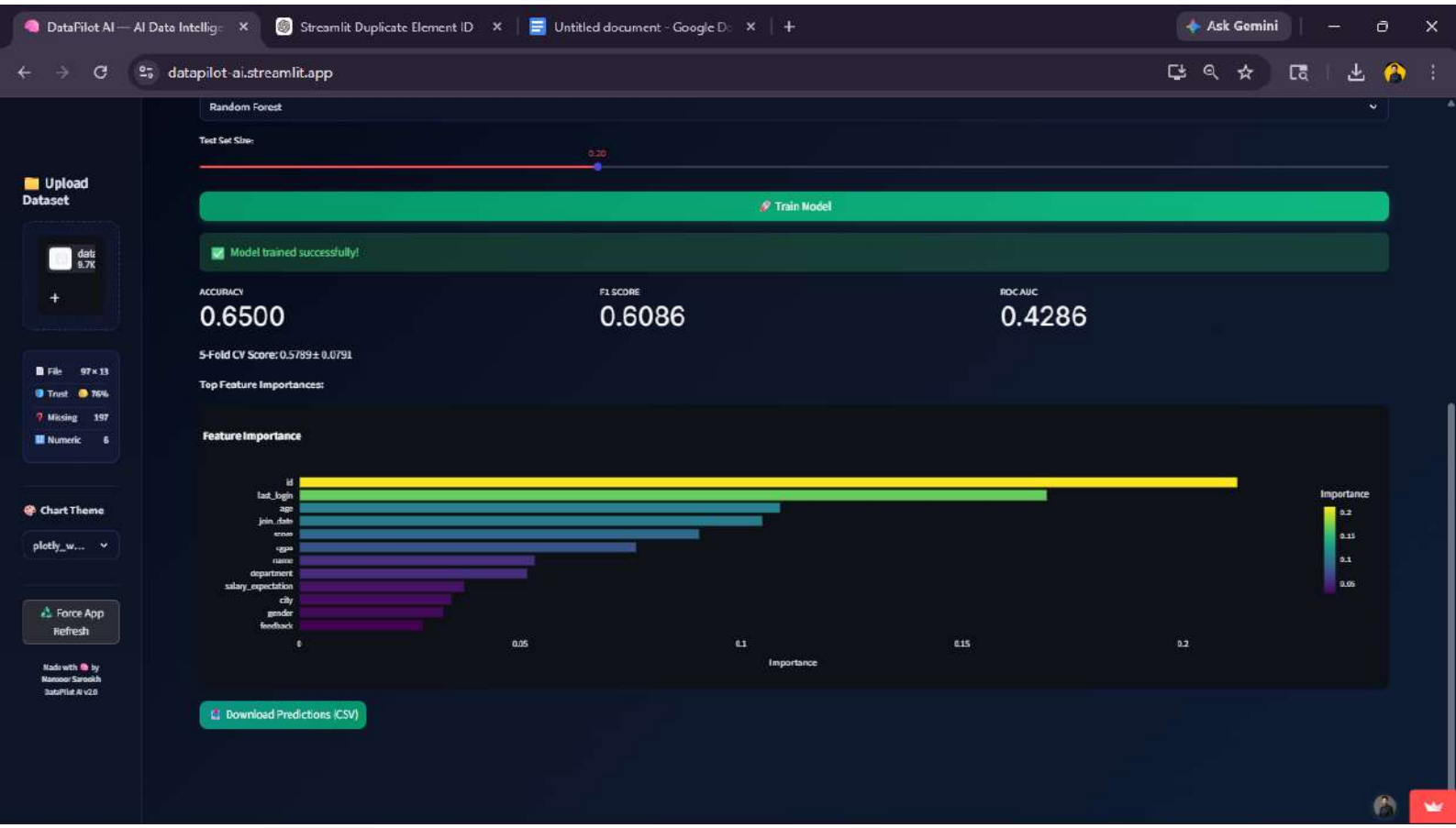
Logistic Regression

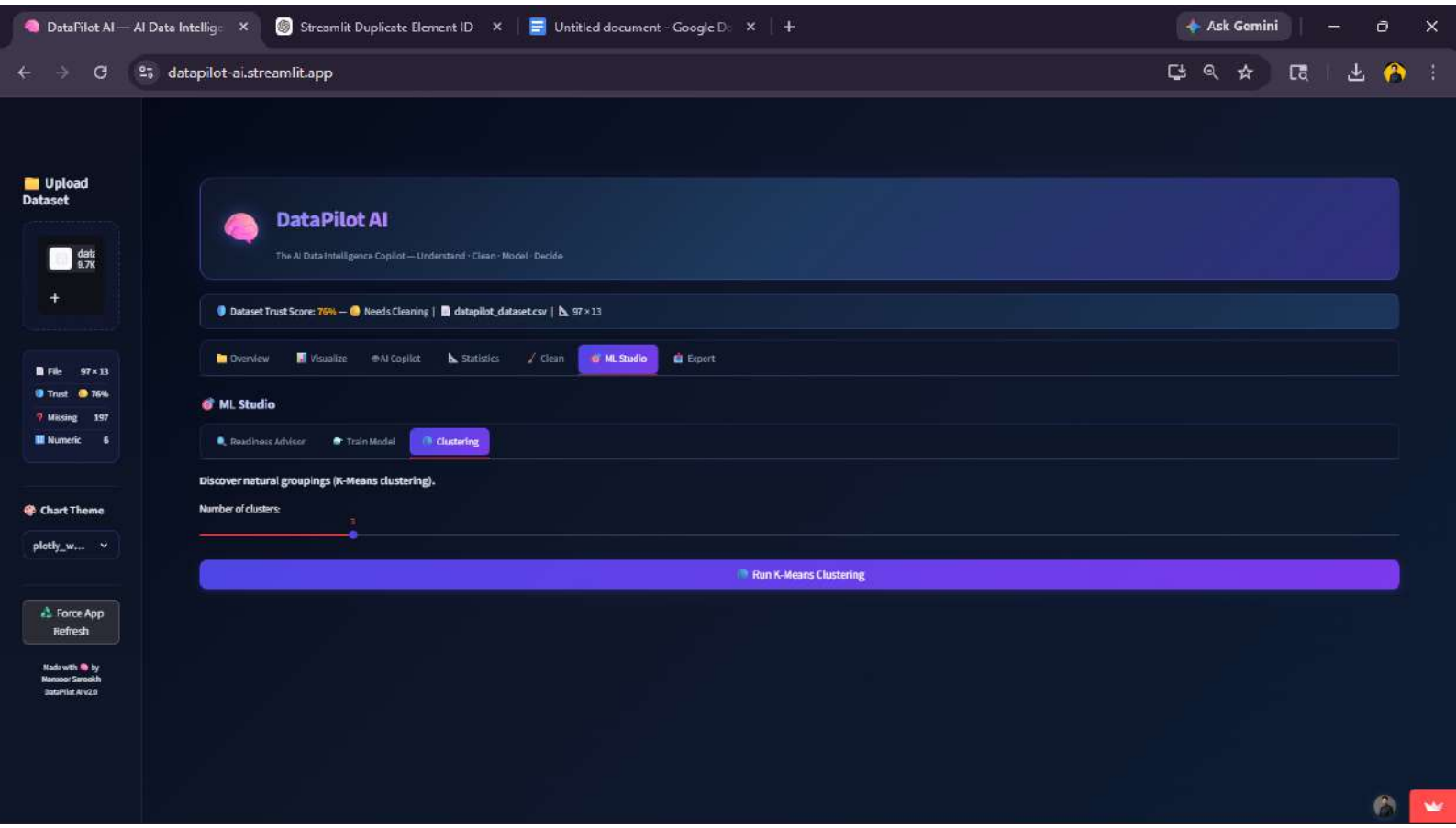
Test Set Size:

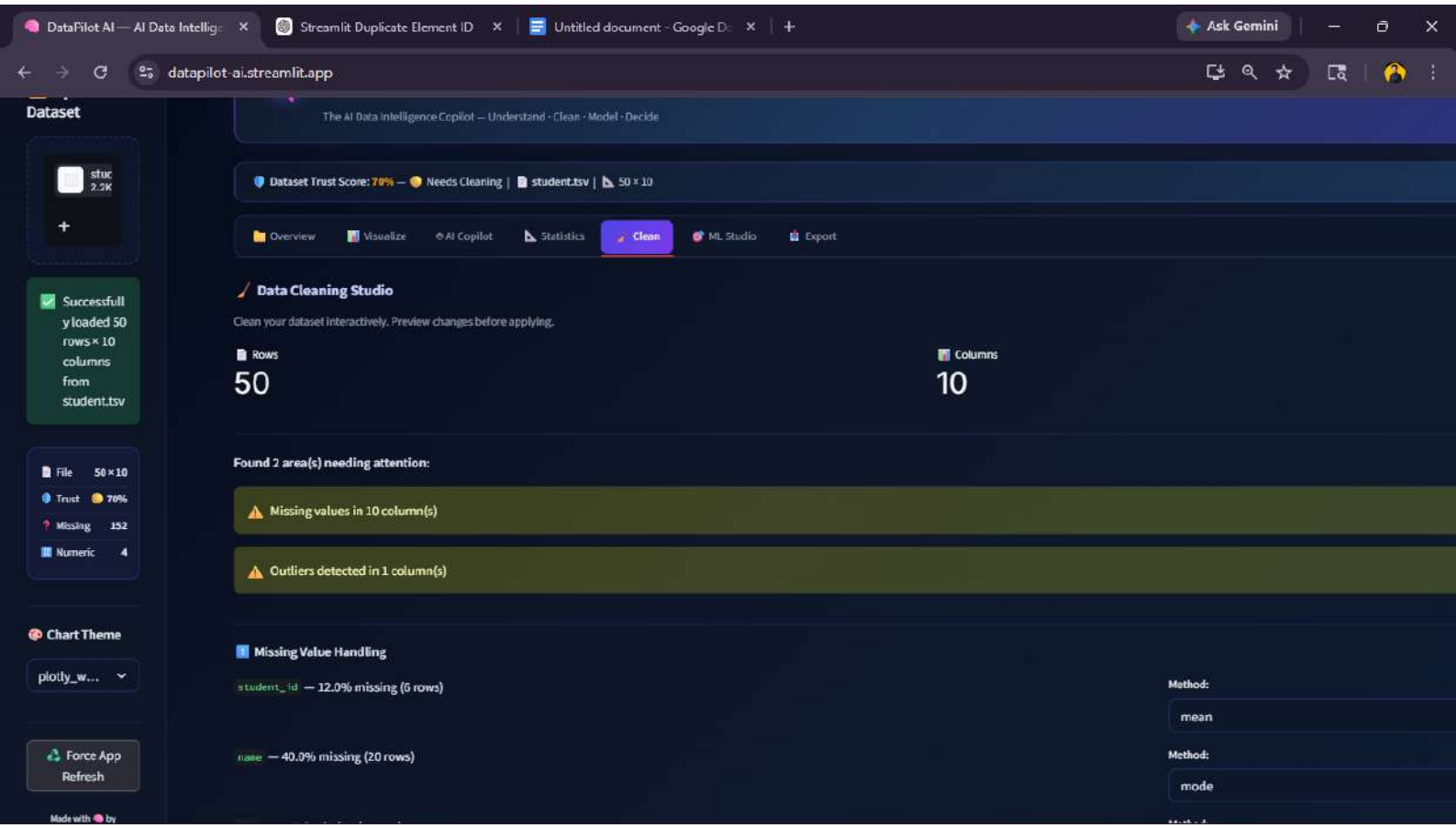
0.20

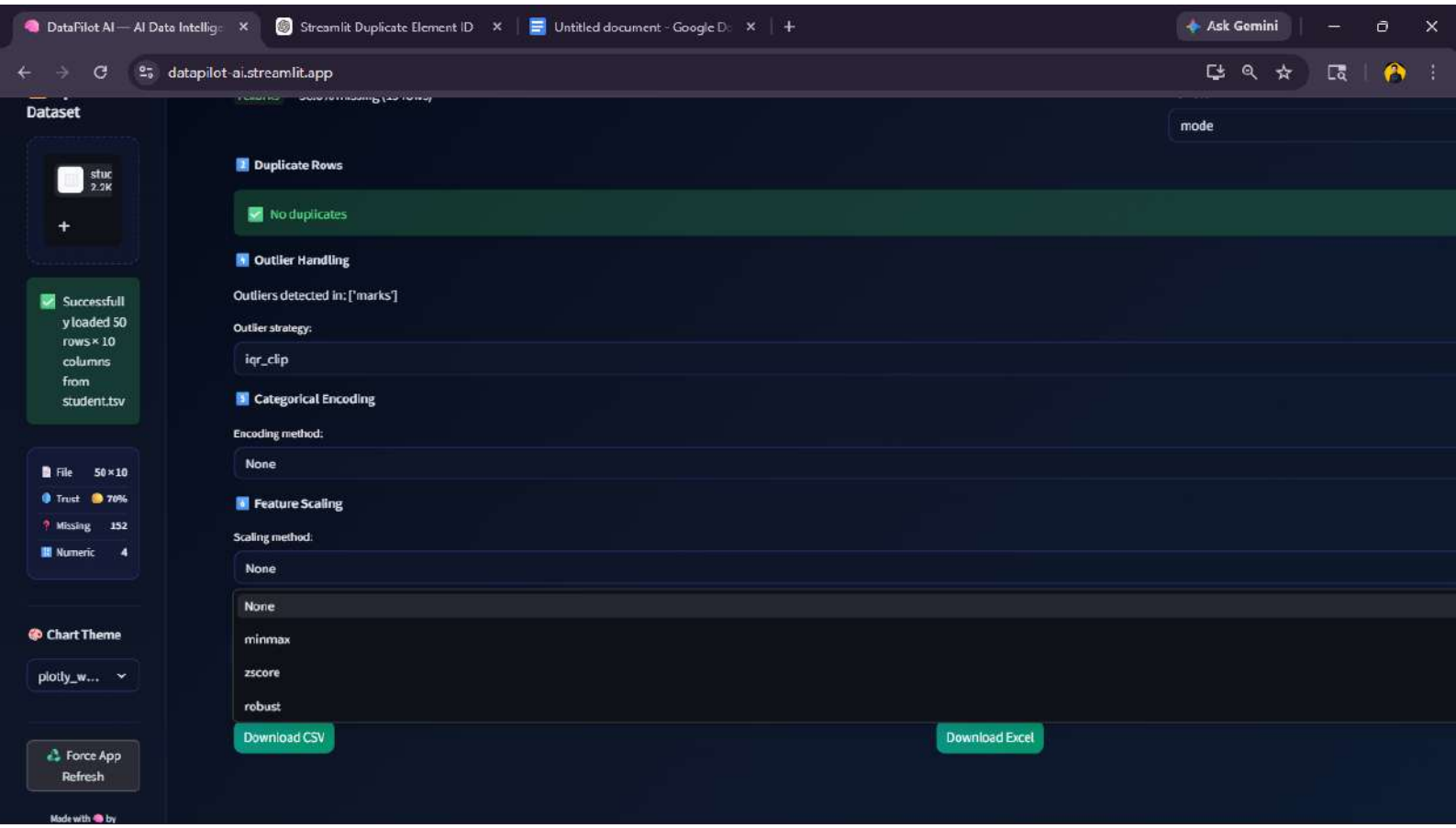
Train Model

Training failed: Unknown label type: continuous. Maybe you are trying to fit a classifier, which expects discrete classes on a regression target with continuous values.


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Random Forest
Test Set Size:
0.20
Upload Dataset
Train Model
Model trained successfully!
ACCURACY
0.6500
F1 SCORE
0.6086
ROC AUC
0.4286
5-Fold CV Score: 0.5789 ± 0.0791
Top Feature Importances:
Feature Importance
id
last_login
age
join_date
score
cgpa
name
department
salary_expectation
city
gender
feedback
0
0.05
0.1
0.15
0.2
Importance
Importance
0.2
0.15
0.1
0.05
File 97 × 13
Trust 76%
Missing 197
Numeric 6
Chart Theme
plotly_w...
Force App Refresh
Made with by Nameer Sarookh DataPilot AI v2.0
Download Predictions (CSV)


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
File 97 × 13
Trust 76%
Missing 197
Numeric 6
Chart Theme
plotly_w...
Force App Refresh
Made with by Nanoor Sarokh
DataPilot AI v2.0
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 76% — Needs Cleaning | datapilot_dataset.csv | 97 × 13
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
ML Studio
Readiness Advisor
Train Model
Clustering
Discover natural groupings (K-Means clustering).
Number of clusters:
3
Run K-Means Clustering


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Dataset
stuc
2.2K
Successfully loaded 50 rows × 10 columns from student.tsv
File 50 × 10
Trust 70%
Missing 152
Numeric 4
Chart Theme
plotly_w...
Force App Refresh
Made with by
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 70% — Needs Cleaning | student.tsv | 50 × 10
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Data Cleaning Studio
Clean your dataset interactively. Preview changes before applying.
Rows
50
Columns
10
Found 2 area(s) needing attention:
Missing values in 10 column(s)
Outliers detected in 1 column(s)
Missing Value Handling
student_id — 12.0% missing (6 rows)
Method:
mean
name — 40.0% missing (20 rows)
Method:
mode


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Dataset
stuc
2.2K
Successfully loaded 50 rows × 10 columns from student.tsv
File 50 × 10
Trust 70%
Missing 152
Numeric 4
Chart Theme
plotly_w...
Force App Refresh
Made with by
mode
Duplicate Rows
No duplicates
Outlier Handling
Outliers detected in: ['marks']
Outlier strategy:
iqr_clip
Categorical Encoding
Encoding method:
None
Feature Scaling
Scaling method:
None
None
minmax
zscore
robust
Download CSV
Download Excel

DataPilot AI — AI Data Intellig | Streamlit Duplicate Element ID | Untitled document - Google D

Ask Gemini

datapilot-ai.streamlit.app

Dataset

stuc
2.2K

+

File 50 × 10
Trust 70%
Missing 152
Numeric 4

Chart Theme

plotly_w...

Force App Refresh

Made with by Manasar Saroakh
DataPilot AI v2.0

Preview Cleaned Data

Before vs After:

Original Rows
50

Original Cols
10

Cleaned Rows
50
↑ 0

Cleaned Cols
10
↑ 0

Operations Applied

	student_id	name	age	gender	email	marks	attendance	city	enrollment_year
0	1	Ayesha	18	Other	test@email.com	90	75	None	
1	None	John	21	None	invalid@	5	None	Peshawar	
2	3	None	None	None	None	80	75	Lahore	
3	None	John	20	None	test@email.com	5	None	Lahore	
4	5	Sara	None	None	None	5	75	Swabi	
5	6	None	19	Male	None	5	None	None	
6	7	Sara	20	Male	invalid@	125	80	Peshawar	
7	8	Sara	21	None	None	80	80	Lahore	
8	9	Sara	18	Male	None	125	None	Lahore	
9	None	John	None	Female	None	60	85	Lahore	

Export Dataset

DataPilot AI — AI Data Intellig | Streamlit Duplicate Element ID | Untitled document - Google D | Ask Gemini

datapilot-ai.streamlit.app

Upload Dataset

data 9.7K

+

File 97 × 13
Trust 76%
Missing 197
Numeric 6

Chart Theme

plotly_w...

Force App Refresh

Made with by Nanoor Sarokh DataPilot AI v2.0

salary_expectation

Assess ML Readiness

55/100

ML Readiness Score

REGRESSION

No duplicate rows

15.6% missing values — significant imputation needed

High-cardinality categoricals: ['join_date', 'last_login'] — use target/hash encoding

7 categorical column(s) need encoding: ['name', 'gender', 'join_date', 'last_login', 'department']

Columns with large scale variance (suggest normalization): ['salary_expectation']

Very small dataset (97 rows) — results may be unreliable

Feature Engineering Suggestions:

[High] DateTime Parsing — `join_date`

[High] DateTime Parsing — `last_login`

[Medium] One-Hot Encoding — `name`

[Medium] One-Hot Encoding — `gender`

DataPilot AI — AI Data Intellig | Streamlit Duplicate Element ID | Untitled document - Google D | Ask Gemini

datapilot-ai.streamlit.app

Upload Dataset

data 9.7K

+

File 97 × 13
Trust 76%
Missing 197
Numeric 6

Chart Theme

plotly_w...

Force App Refresh

Made with by Nanoor Sarokh
DataPilot AI v2.0

DataPilot AI

The AI Data Intelligence Copilot — Understand · Clean · Model · Decide

Dataset Trust Score: 76% — Needs Cleaning | datapilot_dataset.csv | 97 × 13

Overview | Visualize | AI Copilot | Statistics | Clean | ML Studio | Export

ML Studio

Readiness Advisor | Train Model | Clustering

Train an ML model directly in your browser.

Target Column:

id

Detected Problem Type: Regression (97 unique values)

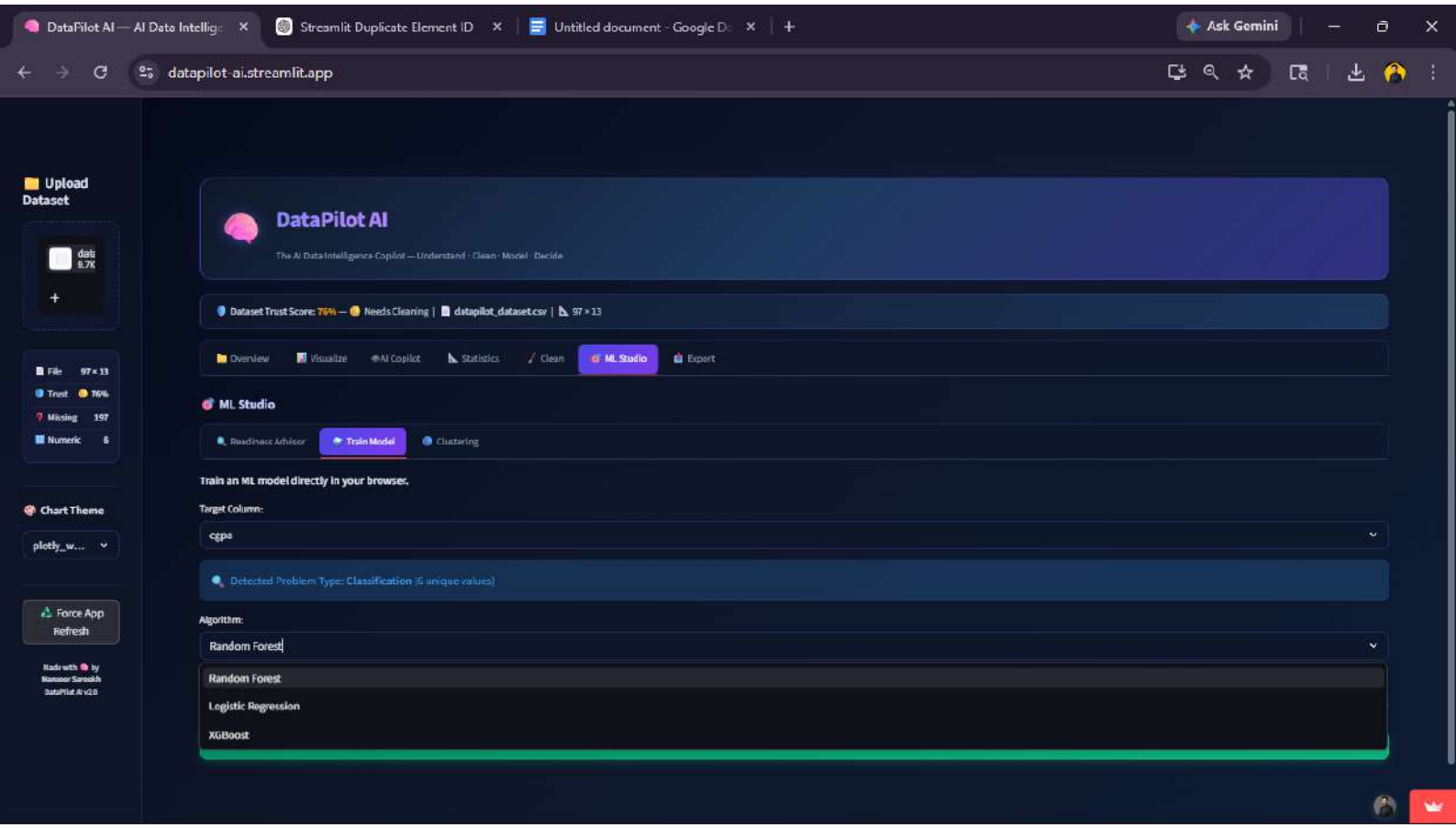
Algorithm:

Random Forest

Test Set Size:

0.20

Train Model


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 76% — Needs Cleaning | datapilot_dataset.csv | 97 × 13
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
File 97 × 13
Trust 76%
Missing 197
Numeric 6
ML Studio
Readiness Advisor
Train Model
Clustering
Chart Theme
plotly_w...
Train an ML model directly in your browser.
Target Column:
cgpa
Force App Refresh
Detected Problem Type: Classification (6 unique values)
Algorithm:
Random Forest
Random Forest
Logistic Regression
XGBoost

DataPilot AI — AI Data Intellig | Streamlit Duplicate Element ID | Untitled document - Google D

Ask Gemini

datapilot-ai.streamlit.app

DataPilot AI

The AI Data Intelligence Copilot — Understand · Clean · Model · Decide

Upload Dataset

data 8.7K

+

File 97 × 13

Trust 76%

Missing 197

Numeric 6

Chart Theme

plotly_w...

Force App Refresh

Made with by Nanoor Sarokh DataPilot AI v2.0

Dataset Trust Score: 76% — Needs Cleaning | datapilot_dataset.csv | 97 × 13

Overview | Visualize | AI Copilot | Statistics | Clean | ML Studio | Export

Export & Report Center

Executive Report | Jupyter Notebook | Data Export

Generate a one-click professional analysis report.

✨ Include AI-generated executive summary (requires Groq API)

Add Custom Insights (Optional)

Insight 1:

e.g. Q3 revenue dropped due to seasonal factors

Insight 2:

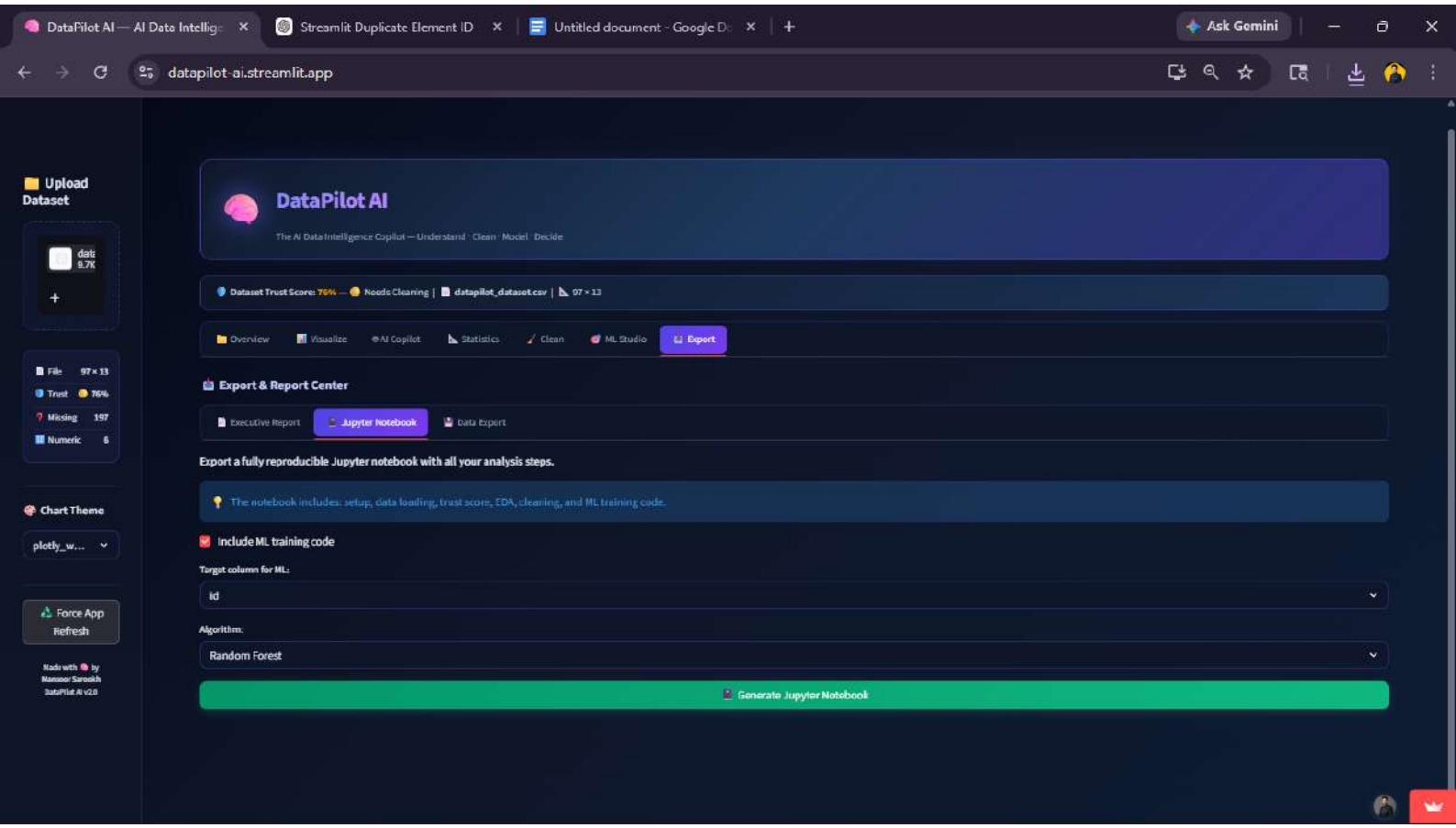
e.g. Q3 revenue dropped due to seasonal factors

Insight 3:

e.g. Q3 revenue dropped due to seasonal factors

Generate PDF Report | Generate HTML Report

Preview AI Summary


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
data 8.7K
File 97 × 13
Trust 76%
Missing 197
Numeric 6
Chart Theme
plotly_w...
Force App Refresh
Made with by Nanoor Sarokh DataPilot AI v2.0
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 76% — Needs Cleaning | datapilot_dataset.csv | 97 × 13
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Export & Report Center
Executive Report
Jupyter Notebook
Data Export
Export a fully reproducible Jupyter notebook with all your analysis steps.
The notebook includes: setup, data loading, trust score, EDA, cleaning, and ML training code.
Include ML training code
Target column for ML:
id
Algorithm:
Random Forest
Generate Jupyter Notebook

DataPilot AI — AI Data Intellig | Streamlit Duplicate Element ID | Untitled document - Google D | Ask Gemini

datapilot-ai.streamlit.app

DataPilot AI

The AI Data Intelligence Copilot — Understand · Clean · Model · Decide

Upload Dataset

data 9.7K

+

File 97 × 13
Trust 76%
Missing 197
Numeric 6

Chart Theme

plotly_w...

Force App Refresh

Made with by Nanoor Sarokh
DataPilot AI v2.0

Dataset Trust Score: 76% — Needs Cleaning | datapilot_dataset.csv | 97 × 13

Overview | Visualize | AI Copilot | Statistics | Clean | ML Studio | Export

Export & Report Center

Executive Report | Jupyter Notebook | Data Export

Export a fully reproducible Jupyter notebook with all your analysis steps.

The notebook includes: setup, data loading, trust score, EDA, cleaning, and ML training code.

Include ML training code

Target column for ML:

attendance

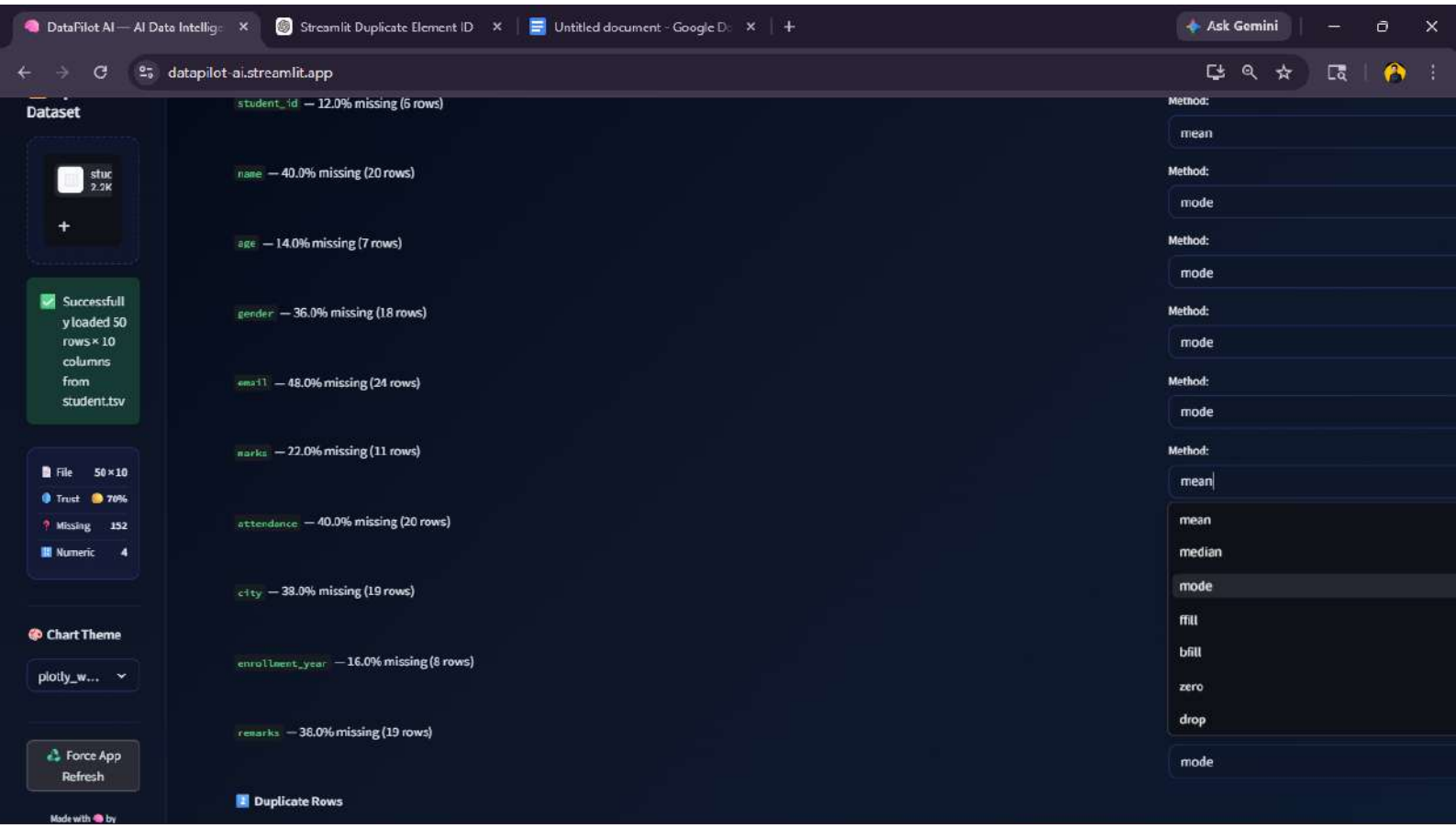
Algorithm:

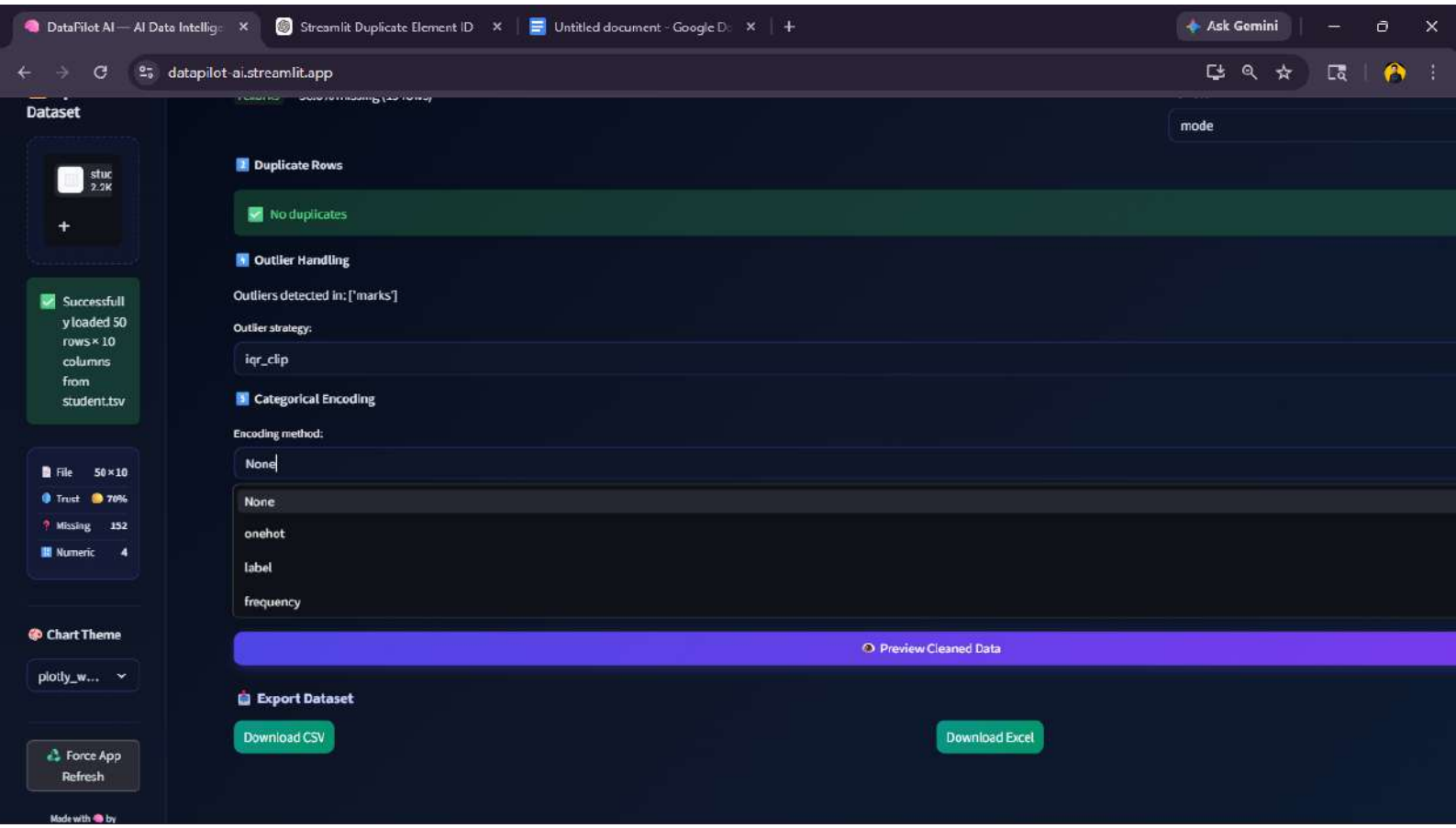
Logistic Regression

Generate Jupyter Notebook

Download .ipynb Notebook

Notebook generated! Open with Jupyter Lab or VS Code.


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Dataset
stuc
2.2K
Successfully loaded 50 rows × 10 columns from student.tsv
File 50 × 10
Trust 70%
Missing 152
Numeric 4
Chart Theme
plotly_w...
Force App Refresh
Made with by
student_id — 12.0% missing (6 rows)
name — 40.0% missing (20 rows)
age — 14.0% missing (7 rows)
gender — 36.0% missing (18 rows)
email — 48.0% missing (24 rows)
marks — 22.0% missing (11 rows)
attendance — 40.0% missing (20 rows)
city — 38.0% missing (19 rows)
enrollment_year — 16.0% missing (8 rows)
remarks — 38.0% missing (19 rows)
Duplicate Rows
Method:
mean
Method:
mode
Method:
mode
Method:
mode
Method:
mode
Method:
mean
mean
median
mode
ffill
bfill
zero
drop
mode


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Dataset
stuc
2.2K
Successfully loaded 50 rows × 10 columns from student.tsv
File 50 × 10
Trust 70%
Missing 152
Numeric 4
Chart Theme
plotly_w...
Force App Refresh
Made with by
mode
Duplicate Rows
No duplicates
Outlier Handling
Outliers detected in: ['marks']
Outlier strategy:
iqr_clip
Categorical Encoding
Encoding method:
None
None
onehot
label
frequency
Preview Cleaned Data
Export Dataset
Download CSV
Download Excel

DataPilot AI — AI Data Intellig | Streamlit Duplicate Element ID | Untitled document - Google D

datapilot-ai.streamlit.app

Dataset

inte 9.2K

+

File 100 × 15

Trust 73%

Missing 343

Numeric 6

Chart Theme

plotly_w...

Force App Refresh

Made with by Mansoor Saroakh DataPilot AI v2.0

Scaling method:

None

Preview Cleaned Data

Before vs After:

Original Rows
100

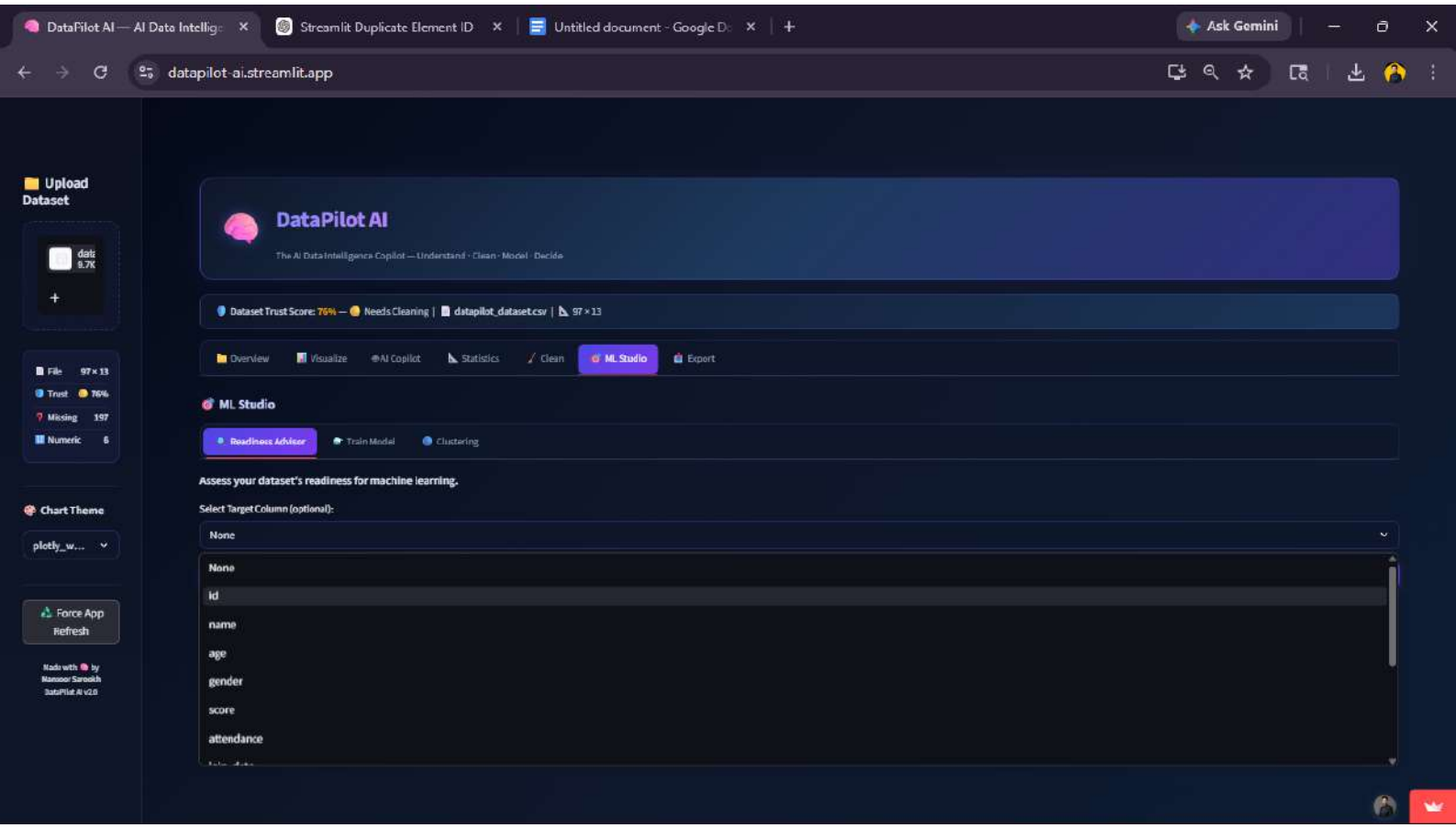
Original Cols
15

Cleaned Rows
97
↓ -3

Cleaned Cols
13
↓ -2

Operations Applied

	id	name	age	gender	score	attendance	join_date	last_login	department	cgpa	feedback	salary_expect
0	1	None	14.5	Other	71.3057	None	2023-08-03 00:00:00	2026-03-21 06:44:42	SE	3.5	Good	
1	2	Ali	21	Female	85.5149	1	2023-11-10 00:00:00	2026-04-03 13:44:42	IT	2.5	None	
2	3	Ali	22	Female	71.3764	1	2023-02-14 00:00:00	2026-03-18 08:44:42	None	None	Excellent	
3	4	None	14.5	None	40.1865	1	2021-03-25 00:00:00	2026-04-24 09:44:42	AI	4	Good	
4	5	John	20	Female	66.7049	None	2023-11-03 00:00:00	2026-04-24 02:44:42	IT	3	Average	
6	7	Sara	22	Female	92.1684	1	2020-04-24 00:00:00	2026-04-07 00:44:42	SE	3	Average	
7	8	Sara	22	None	62.2259	1	2021-04-14 00:00:00	2026-03-23 22:44:42	CS	6.25	None	
8	9	None	14.5	Female	57.8726	0	2020-03-06 00:00:00	2026-04-24 19:44:42	SE	3.5	None	
9	10	Ali	19	Other	62.4736	1	2021-10-08 00:00:00	2026-03-26 07:44:42	IT	6.25	Average	


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
data 9.7K
File 97 × 13
Trust 76%
Missing 197
Numeric 6
Chart Theme
plotly_w...
Force App Refresh
Made with by Nanoor Sarokh DataPilot AI v2.0
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 76% — Needs Cleaning | datapilot_dataset.csv | 97 × 13
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
ML Studio
Readiness Advisor
Train Model
Clustering
Assess your dataset's readiness for machine learning.
Select Target Column (optional):
None
None
id
name
age
gender
score
attendance

DataPilot AI — AI Data Intellig | Streamlit Duplicate Element ID | Untitled document - Google D | Ask Gemini

datapilot-ai.streamlit.app

Upload Dataset

data 9.7K

+

File 97 × 13
Trust 76%
Missing 197
Numeric 6

Chart Theme

plotly_w...

Force App Refresh

Made with by Nanoor Sarokh DataPilot AI v2.0

High cardinality categoricals: ['join_date', 'last_login'] — use target/hash encoding

7 categorical column(s) need encoding: ['name', 'gender', 'join_date', 'last_login', 'department']

Columns with large scale variance (suggest normalization): ['salary_expectation']

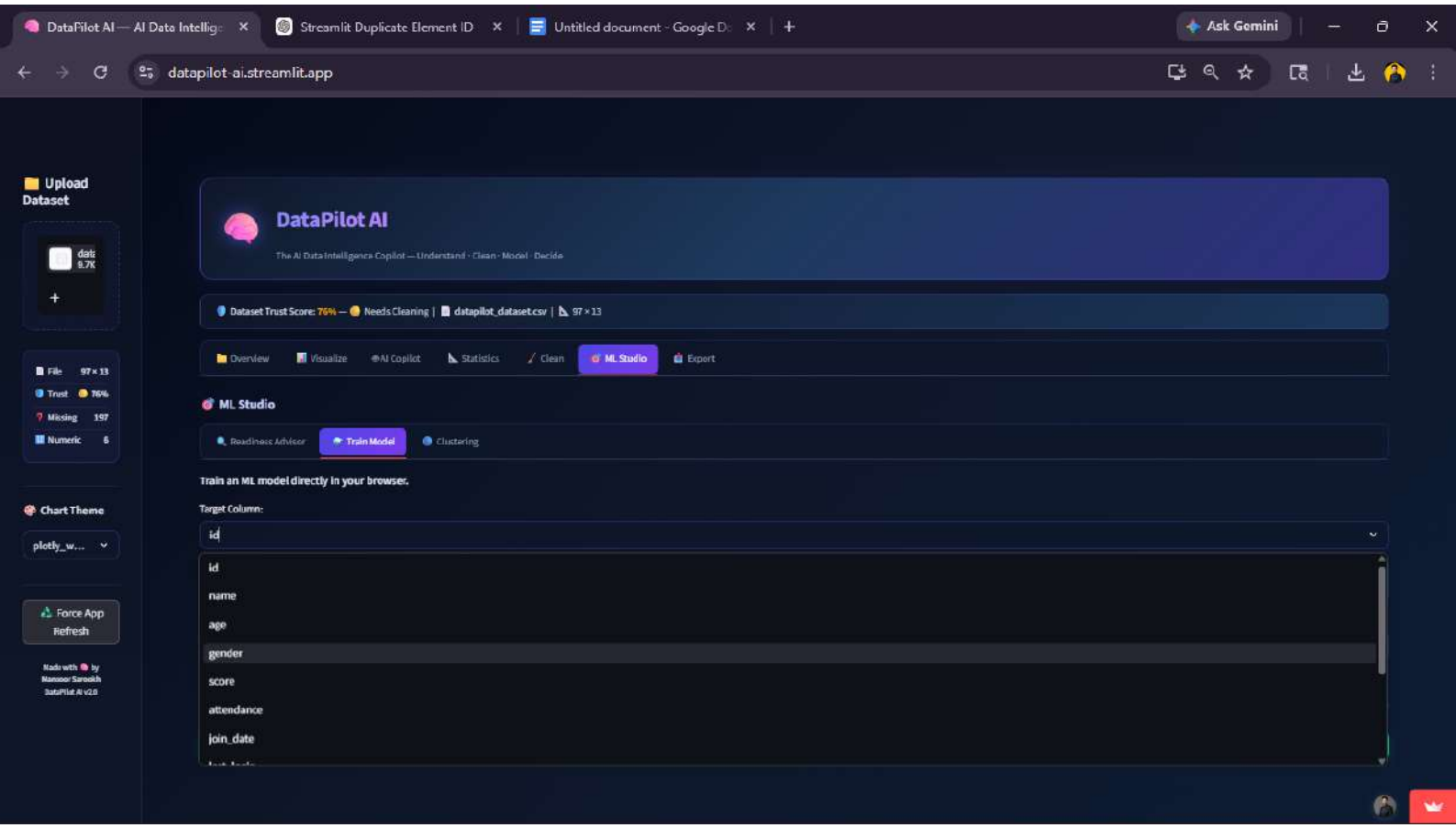
Very small dataset (97 rows) — results may be unreliable

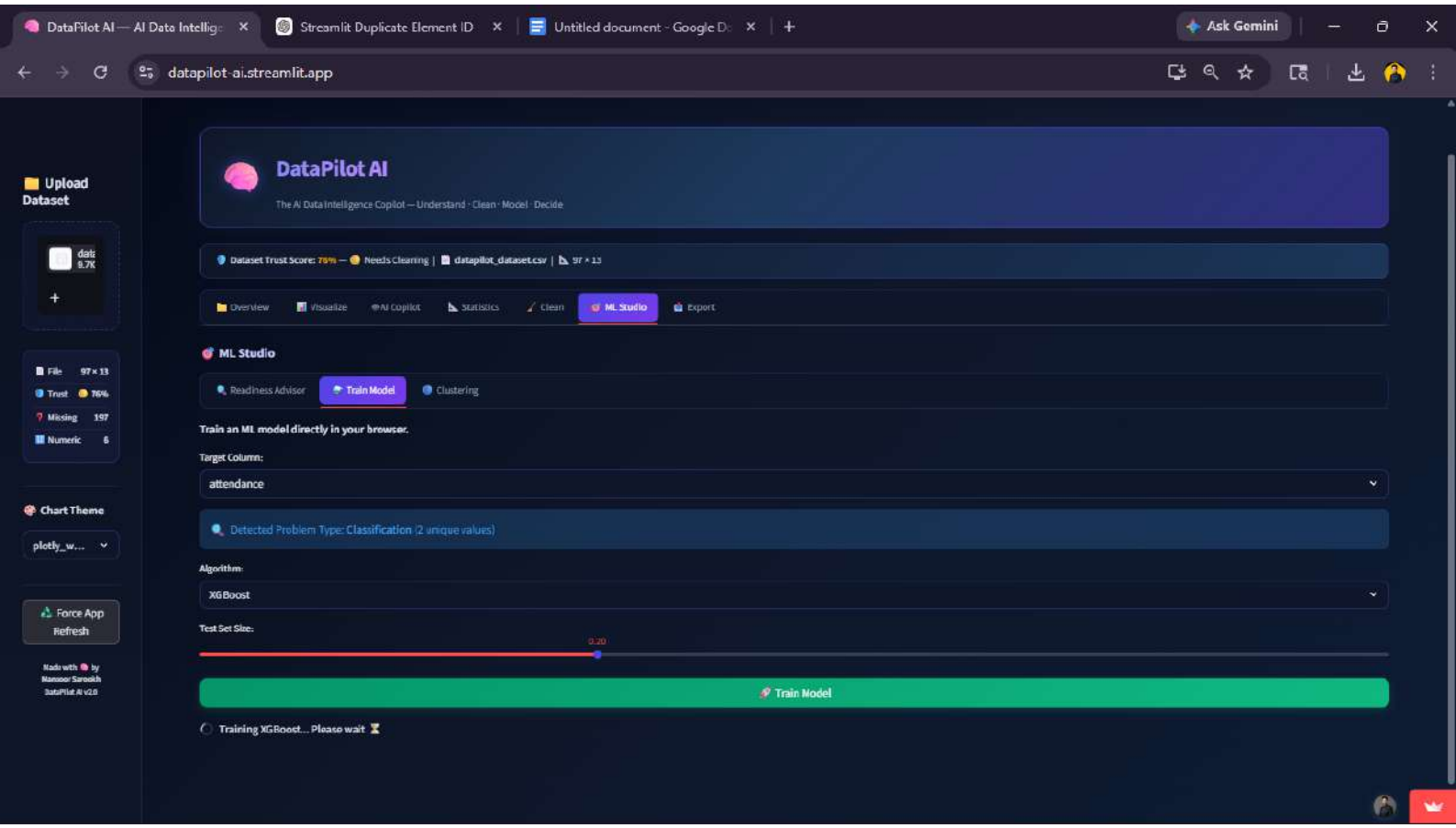
Feature Engineering Suggestions:

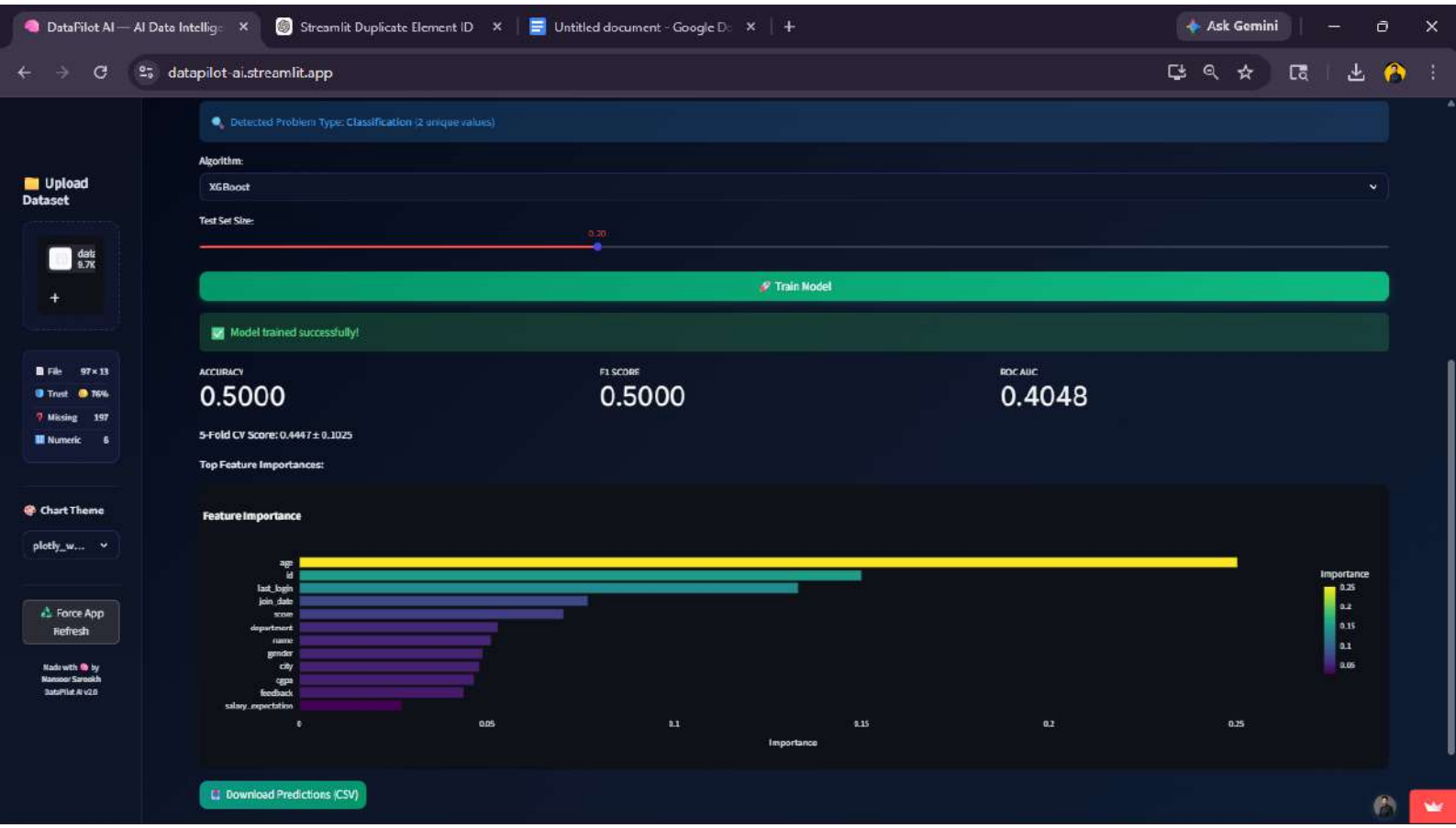
- [High] DateTime Parsing — `join_date`
- [High] DateTime Parsing — `last_login`
- [Medium] One-Hot Encoding — `name`
- [Medium] One-Hot Encoding — `gender`
- [Medium] Frequency Encoding — `join_date`
- [Medium] Frequency Encoding — `last_login`

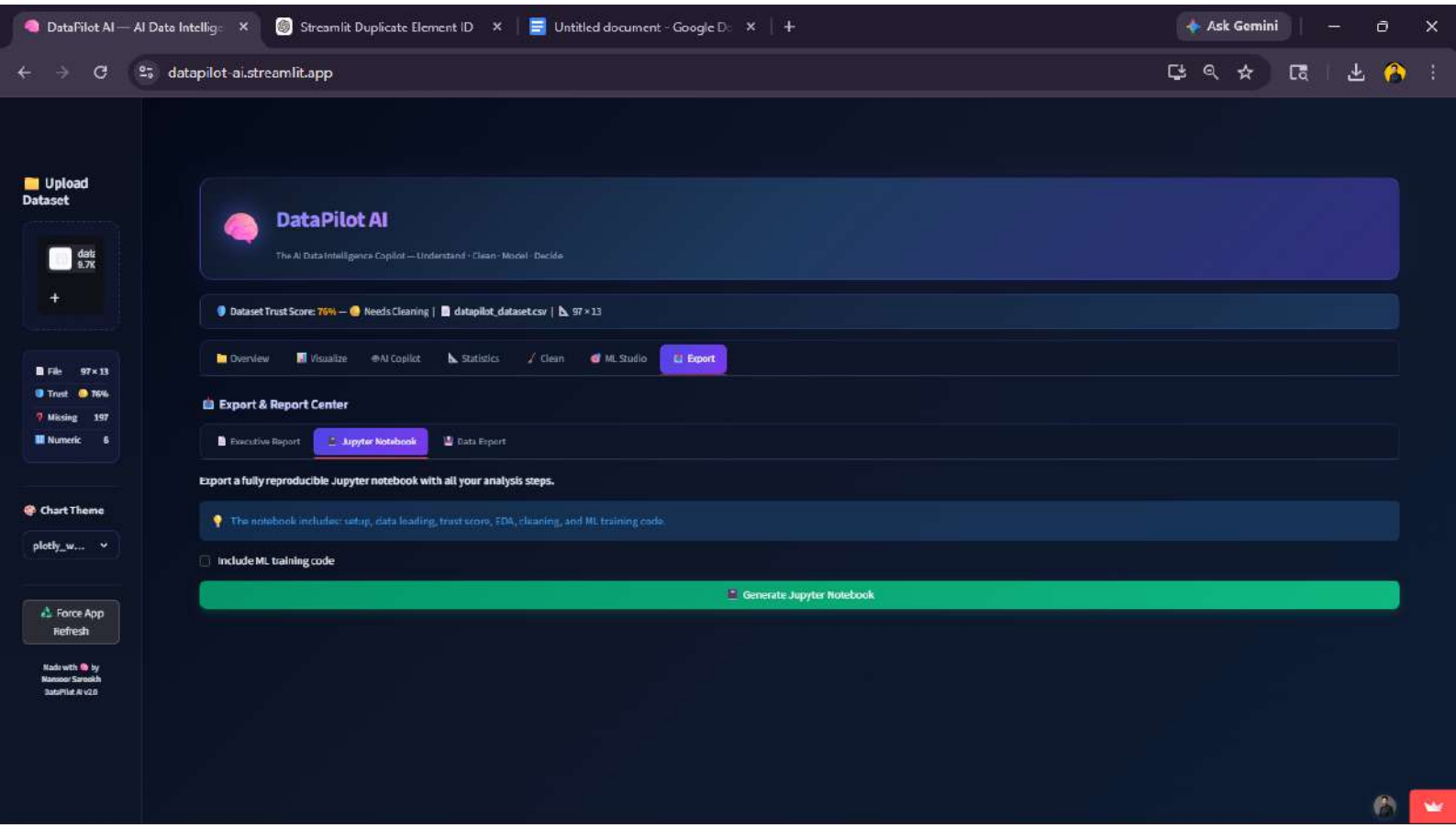
Recommended Algorithms: Linear Regression, Random Forest, XGBoost

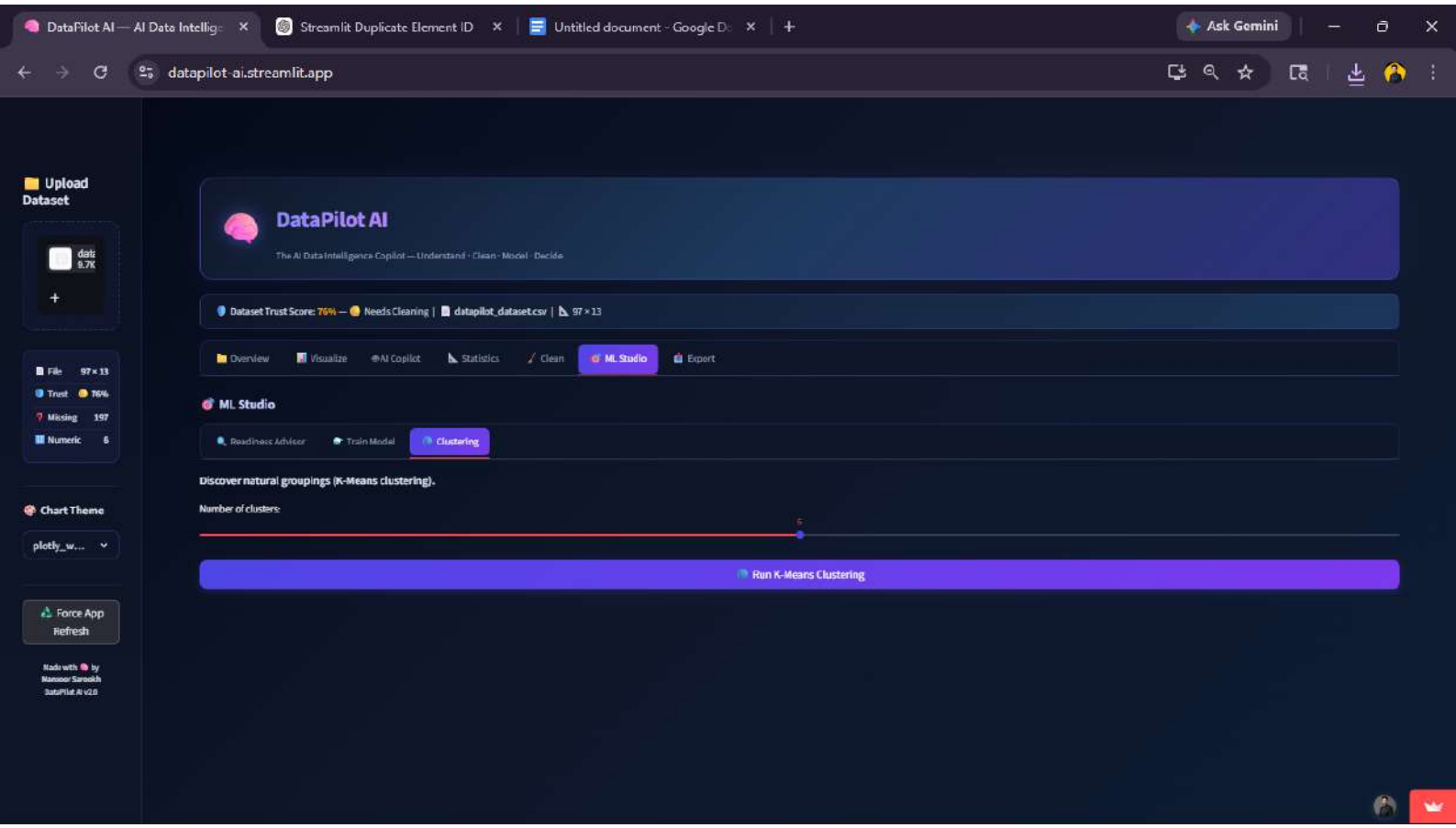
- **Metric:** R^2 / RMSE
- **Expected R2 Range:** 0.60 – 0.85


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 76% — Needs Cleaning | datapilot_dataset.csv | 97 × 13
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
File 97 × 13
Trust 76%
Missing 197
Numeric 6
ML Studio
Readiness Advisor
Train Model
Clustering
Chart Theme
plotly_w...
Train an ML model directly in your browser.
Target Column:
id
id
name
age
gender
score
attendance
join_date
Force App Refresh
Made with by Nanoor Sarokh
DataPilot AI v2.0


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
data 9.7K
File 97 × 13
Trust 76%
Missing 197
Numeric 6
Chart Theme
plotly_w...
Force App Refresh
Made with by Nanoor Sarokh DataPilot AI v2.0
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 76% — Needs Cleaning | datapilot_dataset.csv | 97 × 13
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
ML Studio
Readiness Advisor
Train Model
Clustering
Train an ML model directly in your browser.
Target Column:
attendance
Detected Problem Type: Classification (2 unique values)
Algorithm:
XGBoost
Test Set Size:
0.20
Train Model
Training XGBoost... Please wait


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
File 97 × 13
Trust 76%
Missing 197
Numeric 6
Chart Theme
plotly_w...
Force App Refresh
Made with by Nanoor Sarokh DataPilot AI v2.0
Detected Problem Type: Classification (2 unique values)
Algorithm:
XGBoost
Test Set Size:
0.20
Train Model
Model trained successfully!
ACCURACY
0.5000
F1 SCORE
0.5000
ROC AUC
0.4048
5-Fold CV Score: 0.4447 ± 0.1025
Top Feature Importances:
Feature Importance
age
id
last_login
join_date
score
department
name
gender
city
cgpa
feedback
salary_expectation
0
0.05
0.1
0.15
0.2
0.25
Importance
Download Predictions (CSV)


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
File 97 × 13
Trust 76%
Missing 197
Numeric 6
Chart Theme
plotly_w...
Force App Refresh
Made with by Nanoor Sarokh
DataPilot AI v2.0
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 76% — Needs Cleaning | datapilot_dataset.csv | 97 × 13
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
Export & Report Center
Executive Report
Jupyter Notebook
Data Export
Export a fully reproducible Jupyter notebook with all your analysis steps.
The notebook includes: setup, data loading, trust score, EDA, cleaning, and ML training code.
Include ML training code
Generate Jupyter Notebook


DataPilot AI — AI Data Intellig
Streamlit Duplicate Element ID
Untitled document - Google D
Ask Gemini
datapilot-ai.streamlit.app
Upload Dataset
data 8.7K
File 97 × 13
Trust 76%
Missing 197
Numeric 6
Chart Theme
plotly_w...
Force App Refresh
Made with by Nanoor Sarokh DataPilot AI v2.0
DataPilot AI
The AI Data Intelligence Copilot — Understand · Clean · Model · Decide
Dataset Trust Score: 76% — Needs Cleaning | datapilot_dataset.csv | 97 × 13
Overview
Visualize
AI Copilot
Statistics
Clean
ML Studio
Export
ML Studio
Readiness Advisor
Train Model
Clustering
Discover natural groupings (K-Means clustering).
Number of clusters:
6
Run K-Means Clustering