


# Feature Agent Tutorial

1. Today we will walk you through StepFunction's Feature Agent. This agent will create tuned inputs for AI models from your data. Start by uploading the data you wish to transform.


# Feature Selection App

## Step 1: Data Upload and Preview

Choose a CSV or Excel file

 Drag and drop file here  
Limit 200MB per file • CSV, XLSX, JSON, PARQUET

Browse files

 main\_usage\_data.csv 211.7KB

×

✔ Data loaded successfully. Shape: (3506, 13)


### Data Preview

	customer_id	usage_date	product	total_seats	assigned_seats	active_seats	login_days	feature_e
0	CUST079	2023-11-27	PROD01	119	45	22	2	
1	CUST079	2023-11-14	PROD01	119	45	22	20	
2	CUST079	2023-11-26	PROD01	119	45	32	10	
3	CUST051	2023-10-28	PROD09	171	80	21	7	
4	CUST051	2023-10-29	PROD09	171	80	70	9	

Confirm and Continue

2. Next, the Feature Agent will attempt to identify a category to describe your data. The reason it does this is because it will apply category-specific transformations to your data. If your category differs from the one the Agent identified, you can manually select the category of data you have uploaded.

## Step 2: Category Identification

 AI suggested category: **usage**

Is this correct?

- ☐ Select an option
- ☒ Yes
- ☐ No

Confirm AI Suggestion

3. As a continuation of the prior step, the Feature Agent will suggest some mapping fields to help enforce which columns will and will not be used for transformations in the next step. If you wish to modify them, click the manual selection button and map them via a dropdown.

## Step 3: Field Mapping and Classification

### AI Suggested Critical Field Mappings

#### AI Suggested Mappings:

- customer\_id: customer\_id
- date\_mapping: usage\_date
- revenue\_mapping: Not Found
- target: Not Found

How would you like to proceed?

- ☒ Use AI Recommended Mappings
- ☐ Select Columns Manually


Confirm Mappings

4. After categorization and field mapping, we will move onto the individual feature suggestions. The Agent will provide you with different transformations it will apply to as a final prep before feature creation. You can review each suggestion and its description one-by-one, or if you don't want to apply a given suggestion, you can always skip it. Finally, you can apply all suggestions without reviewing via the bottom right button.

## Step 4: Feature Suggestions

Progress: 0 of 10 suggestions processed

### Suggestions Overview

 Current: Create a derived feature called `seat_utilization_ratio` by calculating the ratio of `active_seats` to `assigned_seats`. This will provide insight into how effectively the assigned seats are being utilized.

### Current Suggestion

Create a derived feature called `'seat_utilization_ratio'` by calculating the ratio of `'active_seats'` to `'assigned_seats'`. This will provide insight into how effectively the assigned seats are being utilized.

Apply This Suggestion

Skip

Apply Remaining (9)

5. After creating features, the Agent will apply statistical transformations on them to help normalize the created features and remove the features that are not as representative of the data. You will be presented with multiple suggested statistical methods to follow. If you want to overrule them, you can manually select and deselect them via the checkboxes.

## Step 5: Feature Selection Methods

### Feature Selection Methods

#### 🧠 AI Recommended Methods:

- **variance\_threshold** (high priority)
- **quasi\_constant\_removal** (high priority)
- **pearson\_correlation** (medium priority)
- **spearman\_correlation** (medium priority)
- **unique\_value\_ratio** (medium priority)
- **low\_variance\_categories** (medium priority)

Select methods to use for feature selection:

- ☒ **variance\_threshold** (high priority) 🧠 AI Recommended
- This method will help eliminate features with low variance, which is important given the presence of features with low unique ratios and potential constant values.
- ☒ **quasi\_constant\_removal** (high priority) 🧠 AI Recommended
- This method is useful for identifying and removing features where one value dominates, which is relevant for features with low unique ratios.
- ☒ **pearson\_correlation** (medium priority) 🧠 AI Recommended
- This method can identify highly correlated numerical features, helping to reduce redundancy in the dataset.

6. The Agent will now present you with the features it created, and it will recommend which ones you should choose. The features are organized by name, what they represent, and their viability for use in AI models. By default, the Agent will recommend features with the Green designation (high benefit for models). Features with the Yellow designation indicate that they may have some viability while the Red designation indicates that the feature is not recommended for use. You can always modify the list based on your preferences via the checkboxes.

## Step 6: Feature Selection

### Feature Recommendations

**Selection Summary:** Selected features based on high correlation, predictive power, and business relevance while avoiding redundancy and low variance.

Selected 7 out of 20 features

<input checked="" type="checkbox"/>	api_calls_per_hour	High correlation; indicates user engagement and product usage.	GREEN
<input checked="" type="checkbox"/>	assigned_seats	High correlation with active seats; significant predictive power.	GREEN
<input checked="" type="checkbox"/>	active_seats	Strong correlation with assigned seats; crucial for understanding usage.	GREEN
<input checked="" type="checkbox"/>	total_feature_usage	Very high correlation; crucial for understanding overall engagement.	GREEN
<input checked="" type="checkbox"/>	feature_a_usage	Very high correlation; strong predictor of engagement and value.	GREEN
<input checked="" type="checkbox"/>	active_hours	High correlation; indicates user engagement and potential retention.	GREEN
<input checked="" type="checkbox"/>	api_calls	High unique ratio and correlation; indicates product usage intensity.	GREEN
<input type="checkbox"/>	average_session_length	Moderate correlation; useful but less impactful than total_feature_usage.	YELLOW
<input type="checkbox"/>	high_usage_day	High correlation; indicates user activity patterns but limited unique values.	YELLOW

- After selecting the features you want to add, the Agent will then apply them onto the data. Download it, and your data will now be ready for use in AI models.

## Step 7: Post Processing

Choose an operation:

- ☒ View Selected Features
- ☐ Download Data
- ☐ Finish

	customer_id	usage_date	assigned_seats	active_seats	feature_a_usage	active_hours	api_
0	CUST079	2023-11-27 00:00:00	0.0898	0.012	0.3013	0.2727	0.
1	CUST079	2023-11-14 00:00:00	0.0898	0.012	0.3243	0.2273	0.
2	CUST079	2023-11-26 00:00:00	0.0898	0.0723	0.5786	0.0909	0.
3	CUST051	2023-10-28 00:00:00	0.2994	0.006	0.8779	0.5455	0.
4	CUST051	2023-10-29 00:00:00	0.2994	0.3012	0.5185	0.1818	0.
5	CUST026	2024-05-01 00:00:00	0.3593	0.006	0.7678	0.0909	0.
6	CUST026	2024-05-06 00:00:00	0.3593	0.2831	0.0851	1	0.
7	CUST026	2024-05-20 00:00:00	0.3593	0.0904	0.1722	0.6364	(
8	CUST096	2023-02-16 00:00:00	0.515	0.4578	0.3894	0	0.
9	CUST096	2023-02-07 00:00:00	0.515	0.5241	0.4204	0.2273	0.