# CODEX MVP Features

This document contains the “key path” scenarios. The intent is to describe the primary user-flow through the application so that completeness can be tested before launch.

1. Pose questions of the graphs so that they can answer things for you

- heat vs. scatter -> histogram

1. “Explain this”

- “Applied classification”

- Take a selection and :

- make decision-tree classifier on this selection vs. everything else

- visualize why it’s different or unique

- “which of the 1000 values matter?”

- LUKAS HAS AN EXAMPLE HE CODED

- also explain *how* the answer was derived

1.1. Filter out garbage

- what’s **good/bad**

- can’t do *anything* without filtering your data set first

- relationships between Features to make sure that it makes sense

- min/max/etc. for a Feature

- derived Features

- mask things out that are garbage

- anything empty/null etc.

- normalization

- manual filtration

- interactive sortable heatmap of values

*1.5. Directed* classification/regression

- don’t optimize hyper-parameters for me

- but GUIDE me

1.5. Train a model to predict X (Classification/Regression)

What did I learn? How well did it work? Cross-validate. Generalized test error…

even on the same features you can use the Classification as a way to LEARN ABOUT THE DATA

1.5. Reliability:

- Error handling so it’s not busted

- Saving sessions

- Selecting Features & Selections and interacting with algorithms and charts

2. “**More like this**”

2. Dimensionality reduction:

- “not minimal”

- high user need because most people won’t do it

- back-end needs this in order to do clustering, but the user doesn’t need it critically right now

3. Real-time algorithms that show each round so you can decide when to stop