High Level Feature List

**Data Cleaning features (ranked by priority)**

Column Specific cleaning

V1. (2 weeks)

1. Clean Null values
2. Ability to rename
3. Ability to drop selecting columns for export
4. Ability to change datatype
5. Some DQ metrics for column hover
6. Save changes to a log ,
7. Ability to Normalise
8. One – Hot encoding
9. Numeric to categorical (Binning)
10. Deal with outliers
11. Special Character removal for text column
12. Trim White Space (trailing/leading)
13. Move column index

V2. 2 weeks

1. Dimensionality reduction
2. Target Variable Selection
3. Preset Features
4. Create new columns

V3.

1. Expand records (Data Expansion)
2. Pivot table

Dataset Specific Features

NB: Note what to do for datasets with columns of more than a certain number

A white board with blue writing

Description automatically generated

Sample UI for Data Cleaning Page

**Data Importing Features**

V1. (1 week)

1. Different CSV formats
2. Excel 1 sheet

V2. (2 weeks)

1. Multiple File imports
2. Remove metadata.
3. Excel Multiple sheets

A white board with writing on it

Description automatically generated

Sample UI for Data Import Page (More research to finetune)

**Data Preview Features**

V1. (1 week)

1. Show sample Data
2. Select Primary Key
3. Select Target Variable

V2. (2 weeks)

1. Solution for 1000 column edge case
2. Dimension reduction

**Data Profiling Features**

V1.

Profiling 🡪 Number of records

Number of fields

File name

Data Types

Data Quality Issues 🡪 Missing Values

Data Quality Scan

Duplicates

Outliers

Cardinality

Feedback loop for visuals

Evaluation (engagements)

Tool tip

V2.

Statistical Analysis 🡪 Distribution plots

Skewness/ Kurtosis

Correlation plot

Chart Annotation

Save Charts

History log

Export Visuals

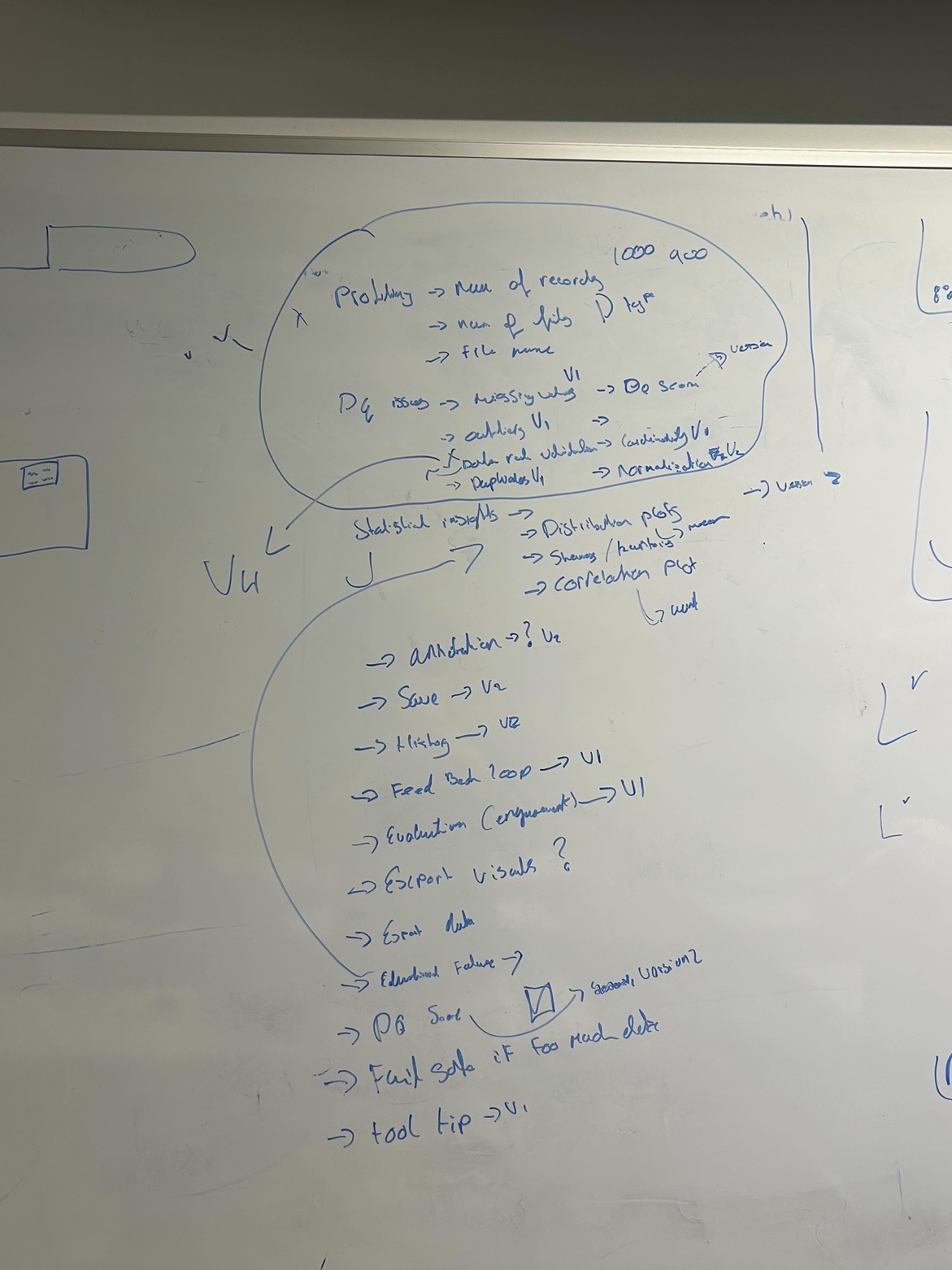
Export data of Visuals

Filters

V3.

Data rule validation

Fail safe to prevent app breaking from too much data



**Data Export Features**

V1.

1. Export to CSV
2. Show sample

V2.

1. Export to Excel
2. Show DQ score
3. Go back to data cleaning/profiling
4. Change file name

V3.

1. Connect to DB
2. Connect to collab
3. Export to Cloud

A white board with writing on it

Description automatically generated

**Timeline:**

Week 5: Project Design and Interim Report

Week 6: Interim Report

Week 7: Data Cleaning V1 and Data Profiling V1

Week 8: Data Preview V1 and Data import V1

Week 9: Data Import V2

Week 10: Data Profiling V2

Week 11: Data Cleaning V2 and testing

Week 12: Bug fixing/Report

Week 13: Report

A whiteboard with writing on it

Description automatically generated

A white board with writing on it

Description automatically generated

Overview Workflow of App showing the various stages of:

Data Import 🡪 Data Preview 🡪 Data Profiling 🡪 Data Cleaning 🡪 Data Export