Here is the complete detailed report of all **9 steps (Step 0–8)** in your CultureBank pipeline, covering **functionality, importance, and column breakdowns**:

## **Step 0: Cultural Relevance Classifier**

### **Functionality**

Filters input comments to determine whether they are culturally relevant. A classifier identifies if a comment pertains to cultural identity, values, or practices.

### **Importance**

Ensures that only meaningful cultural data flows through the pipeline, eliminating irrelevant noise.

### **Key Columns**

* **comment**: The raw user input.
* **is\_culturally\_relevant**: Boolean label for cultural relevance.
* **relevance\_score**: (Optional) Confidence of classification.

## **Step 1: Knowledge Extractor**

### **Functionality**

Extracts structured cultural data (e.g., groups, behaviors, goals) from culturally relevant comments using a language model.

### **Importance**

Transforms unstructured language into a structured cultural taxonomy for processing and analysis.

### **Key Columns**

* **vid / vid\_unique**: IDs to trace data origin.
* **comment\_utc**: Timestamp.
* **cultural group, context**: Identity and setting.
* **actor / recipient**: Who performs and who receives the cultural behavior.
* **relation**: Social connection between actor and recipient.
* **actor's behavior / recipient's behavior**: Actions taken.
* **goal**: Motivation for the action.
* **other descriptions**: Additional narrative content.
* **norm**: Whether the behavior is considered normative (1), non-normative (0), or mixed ([1, 0]).

## **Step 2: Negation Converter**

### **Functionality**

Detects and transforms statements of cultural **absence** (e.g., “We don’t do this”) by modifying extracted descriptors.

### **Importance**

Captures culturally meaningful **exceptions or contrasts** which are often overlooked in typical extraction.

### **Key Column Affected**

* **norm**: Modified to reflect negated behaviors (e.g., becoming 0 instead of 1).

## 

## **Step 3: Clustering**

### **Functionality**

Groups similar extractions into clusters using semantic similarity (e.g., sentence embeddings).

### **Importance**

Identifies consensus and variation in how cultural behaviors are described. Reduces duplication.

### **Key Columns**

* All Step 1 fields
* **cluster\_id**: Unique ID for the cluster.
* **cluster\_size**: Number of descriptors in the cluster.
* **raw\_sample\_vids / raw\_samples**: Track entries in the cluster.
* **raw\_sample\_norms / raw\_sample\_times**: Normative diversity and temporal trends.

## **Step 4: Cluster Summarizer**

### **Functionality**

Generates an LLM-based summary of each cluster to represent shared cultural meaning.

### **Importance**

Creates a **readable, representative descriptor** of each cluster—human- and model-friendly.

### **Key Columns**

* **cluster\_id**
* **topic**: Cluster theme.
* **actor's behavior / goal / recipient\_behavior**: Synthesized actions and motivations.
* Inherits structured fields from Step 3.

## **Step 5: Topic Normalizer**

### **Functionality**

Standardizes cultural group names and topic labels using a normalization mapping.

### **Importance**

Enables search, comparison, and taxonomy alignment across diverse expressions.

### **Key Columns**

* **representative\_cultural group**: Canonical group label.
* **representative\_topic**: Normalized topic.
* **representative\_cultural group\_count / \_cluster\_id**: Metadata for frequency and grouping.

## **Step 6: Agreement Calculator**

### **Functionality**

Scores how much **agreement** exists within each cluster about whether the behavior is a norm.

### **Importance**

Measures community consensus, critical for evaluating representativeness and reliability.

### **Key Columns**

* **norm**: Final agreement score (1.0, 0.0, or NaN).
* **raw\_sample\_norms**: Source values for agreement.

## **Step 7: Content Moderation**

### **Functionality**

Flags clusters for potential **controversial content or PII**, using classifiers and NER (Named Entity Recognition).

### **Importance**

Ensures safe, ethical training data and prepares content for **human review if needed**.

### **Key Columns**

(from output\_for\_annotation.csv)

* **text\_for\_controversial\_prediction**: Input for moderation.
* **pred\_label / pred\_score**: Model output.
* **pii\_result / keywords\_list**: Named entities detected.
* **controversial\_or\_PII**: Human annotation flag.

## **Step 8: Final Formatter**

### **Functionality**

Packages cleaned, moderated cultural knowledge into a final format for model fine-tuning or deployment.

### **Importance**

Consolidates all upstream steps into a consistent, structured, filtered dataset.

### **Key Columns**

* **actor\_behavior / recipient\_behavior / goal / topic / context**: Final descriptors.
* **agreement**: Final agreement score.
* **num\_support\_bin**: Indicates how many examples support the behavior.
* **time\_range**: Year-wise distribution of supporting comments.