# Data Cleaning Project I & II

## Workflow Overview

Anjali Jain Summer 2024

# For Historical Pricing Trend

## Perform Data Cleaning

Remove any duplicate records.
To make sure currency formats are consistent (e.g., all prices in USD).
Fill in or remove missing values, particularly for prices and dates.
Correct inconsistent  date formats.
Normalize dish names where slight variations exist (e.g., "Chicken Gumbo" and "Chicken gumbo" should be treated as the same dish).

## Obtain Clean Data

## Perform Data Analysis

Group data by year and calculate average prices for each dish.
Identify price trends over the years for popular dishes.
Analyze price variations across different locations and events.

## Obtain Pricing Insights

Target (Main) use case U1 Diagram: data cleaning is necessary and sufficient

# ForPredicting Future Trends

## Perform Extensive Data Cleaning

Remove any duplicate records.
To make sure currency formats are consistent (e.g., all prices in USD).
Fill in or remove missing values, particularly for prices and dates.
Handle missing value
Correct any typographical errors in dish names and descriptions.
Normalize dish names where slight variations exist (e.g., "Chicken Gumbo" and "Chicken gumbo" should be treated as the same dish).

## Obtain Clean Data

## Perform Data Analysis

Group data by year and calculate average prices for each dish.
Identify price trends over the years for popular dishes.
Analyze price variations across different locations and events.

## Obtain Pricing Insights

## Perform Extensive Data Cleaning

Gather additional data sources (e.g., economic data, food supply chains, cultural trends).
Integrate external data with the cleaned historical menu dataset.

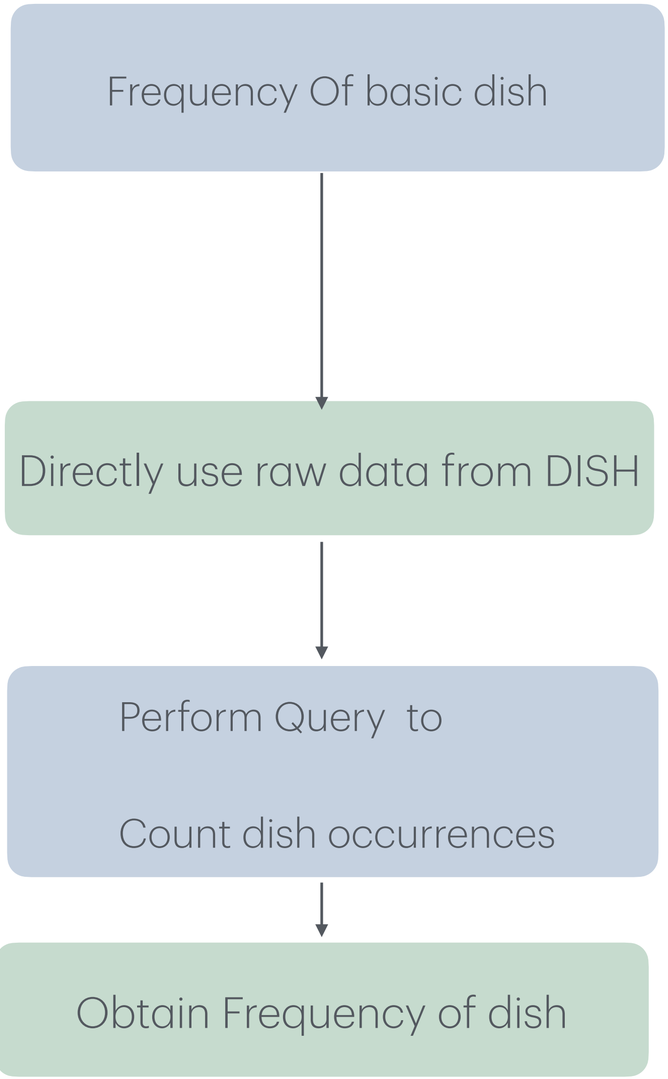
## Perform Predictive Modeling

Use machine learning models to predict future menu trends and pricing.

## Future Trend Prediction

"Never Enough" Use Case U2: Data Cleaning is Not Sufficient

To count the number of times each dish appears across all menus.




Zero Data Cleaning" Use Case U0 diagram: Data Cleaning is Not necessary

# Workflow Overview

Target (Main) use case U1
Diagram: Data cleaning is necessary and sufficient