## **INMT5526 Group Assignment**

Group 27

Angela Jacinto (23778435), Chloe Wu(24019456), Tingwei Hong(24110374), Vinayak Gupta (24066272)

## **Data-Driven Insights for Strategic Expansion**

## I. Executive Summary

This report presents valuable insights derived from a customer survey that analyzes preferences and knowledge of various world cuisines across several demographic segments in the United States. The goal is to provide guidance for a multinational restaurant franchisee as they plan to expand their operations, refine offerings, and design targeted campaigns.

Cuisines such as Italian and Mexican consistently emerged as favorites among respondents, with differences observed across age groups and income levels. These findings present opportunities for the franchisee to enhance customer satisfaction and streamline business strategies. The report also demonstrates how leveraging an interactive dashboard can enable the business to effectively tailor its strategies based on data-driven insights.

## II. Introduction

In today's fast-paced food industry, knowing what consumers prefer is key for businesses, especially those that operate globally. Consumer preferences influence product development, marketing, and decisions about where to expand. For large restaurant chains, understanding what types of food people like helps them create menus that appeal to different cultures and age groups. As competition grows, businesses need to use data to spot trends, offering personalized options that boost customer satisfaction and build brand loyalty.

Consumer preferences are shaped by numerous factors, including psychological, social, and cultural influences, as well as personal tastes and environmental considerations. For instance, convenience, health consciousness, and sustainability have become increasingly significant in shaping food choices, particularly in a post-pandemic world where consumers prioritize transparency, ethical sourcing, and health-conscious options (Open Knowledge, 2020; Martínez-Ruiz, M.P.). Understanding these evolving preferences helps businesses align their offerings with customer expectations while creating targeted marketing strategies for specific demographics.

This report provides a detailed analysis of customer preferences for world cuisines across several demographic segments in the United States, while understanding how factors such as age, income, and region influence food choices. The data analysis was conducted using a comprehensive dashboard that allows cross-filtering between data variables, which guides the viewer in determining actionable insights for strategic decision-making. Furthermore, the report

will discuss the methodology for data collection, processing, and analysis, including key metrics used to derive insights. It will also highlight the design and organization of the dashboard, ensuring that it delivers clear and meaningful visualizations. The findings section will present valuable insights, followed by actionable business recommendations.

#### 2.1 Background and Business Need

In the context of our analysis, understanding consumer preferences is vital for multinational restaurant franchises looking to expand into new markets. These insights allow businesses to identify which cuisines are favored by different demographic groups and regions, providing a foundation for strategic expansion and targeted marketing campaigns. As consumer preferences vary widely across age groups, income levels, and cultural backgrounds, businesses must be equipped with the data needed to make informed decisions. For example, younger demographics might show a stronger preference for fast-casual dining and fusion cuisines, while older demographics may lean toward traditional and comfort foods. (Tastewise (2024))

The need for data-driven decision-making is especially important for companies aiming to enhance their market presence. By understanding which regions show the highest interest in specific world cuisines, franchises can develop localized menus and marketing strategies, ensuring they appeal to regional tastes while maintaining operational efficiency. This strategic approach minimizes risk and maximizes the potential for success in new markets. (Ledue, N. (2021, January 4), (Martínez-Ruiz, M. P).

#### 2.2 Research Question

"How do demographic factors such as age, income, gender, and region influence the preferences and knowledge of world cuisines among U.S. respondents, and which cuisines are most favored across these groups?"

This question serves as a guiding framework, focusing on how demographic factors such as age, income, gender, and region impact respondents' preferences, knowledge, and interest levels in world cuisines, while also identifying the most favored cuisines across these segments.

By answering this question, the analysis aims to give useful insights that will help the franchise in selecting the best places to expand. Additionally, it will also help the franchise develop marketing campaigns that are suited to the cultural preferences of different consumer groups.

#### III. Dataset Overview

#### 3.1 Dataset / Data Model

The dataset contains survey responses about food cuisine preferences from respondents across the United States. Each row represents a single respondent's answers, including their demographic information and ratings for a variety of world cuisines, with ratings ranging from 1 to 5, where 5 is the highest score. Additionally, the dataset captures information about their knowledge and interest in these cuisines.

Before the analysis, the dataset underwent a transformation and cleaning process. Missing or incomplete values were addressed, and certain variables were standardized for consistency. For example, age and household income were grouped into more meaningful ranges to make the data more interpretable. One of the key steps was transforming the dataset from wide format to long format. This allows each respondent to have multiple rows representing their ratings for different cuisines, making the data more efficient to analyze.

## 3.2 Data Dictionary

The data dictionary documents the data assets with relevant context such as data type and description, making it easy to understand and discuss data across teams. In addition, it helps detect anomalies quickly and avoid data inconsistencies (Team Atlan, 2022). We establish a data dictionary according to the new data table after data cleaning.

Table 1. Data Dictionary

Variable	Data Type	Description	Example Values
RespondentID	Text	Unique Identifier of the respondent.	3308895255
Gender	Text	The gender of the respondent.	Male, Female
Age	Text	The age group of the respondent.	Young (18-29 years old), Young Adult (30-44 years old), Middle-aged (45-60 years old), Seniors (over 60 years old)
Household Income	Text	The household income range of the respondents.	\$0 - \$24,999, etc.
Education	Text	The education level of the respondent.	Less than high school, Bachelor's degree
Location	Text	US census region where the respondent is located.	West South Central, Pacific
Cuisine Knowledge Level	Text	The level of knowledge the respondent has in world cuisines.	High, Medium, Low, None

World Cuisines Interest Level	Text	The level of interest the respondent has in world cuisines.	High, Medium, Low, None
Cuisine	Text	The cuisine being rated by the respondent.	Italy, Japan, Mexico, etc.
Rating	Text	The respondent's rating of different cuisines.	1, 2, 3, 4, 5
Rating Category	Text	A categorical representation of the rating.	Low (1-2), Medium (3), High (4-5)

## IV. Methodology

#### 4.1 Data Preprocessing

The provided dataset required significant cleaning and transformation to ensure consistency and accuracy for analysis. Initially, the dataset was in a wide format with numerous columns, making it inefficient to work with. Additionally, several cells contained missing values, and there were issues with inconsistent naming formats and unwanted characters. To address these, the dataset was cleaned by removing rows with insufficient or incomplete data, especially in rows where most cells had no value. Special characters were removed, column names and values were standardized, and certain variables were transformed into more meaningful categories. The key transformation was converting the data from wide format to long format, allowing each respondent to have multiple rows. One for each cuisine rating; making the dataset clearer and more manageable for analysis.

The primary tool used for this process was Power BI Desktop. It enabled the import of the provided dataset in CSV format, transformation of the dataset's structure, and subsequent analysis through the Power Query Editor. The Power Query Editor was key to the data transformation process, allowing for the handling of missing values, data restructuring, and unpivoting. The unpivot function was crucial in converting the dataset from a wide to a long format, making it more suitable for analysis. Additionally, the column type transformation feature ensured the use of appropriate data types by converting text values to integers and vice versa for accurate calculations. Power BI also allowed for the creation of new columns, such as the Rating Category, for grouping.

The cleaning and transformation steps are detailed below:

1. Replacing 'N/A' and Empty Strings with 'No Response'

The dataset had various missing values represented as 'N/A' or empty strings. Replacing these with 'No Response' helps in distinguishing missing data clearly, allowing us to handle incomplete records more effectively during analysis.

#### 2. Removing Special Characters

The dataset contained special characters such as "Ê" in both column names and data values. We remove them to ensure data consistency and readability.

## 3. Renaming Columns

Several column names contained special characters and were in question format. The columns were renamed by removing periods, colons and by applying proper case formatting to ensure clean and proper names to standardize and make the dataset simpler to understand.

#### 4. Unpivoting Cuisine Columns

To make the data more manageable, it was transformed to a long format, with one row per respondent-cuisine pair. This is essential for a more efficient analysis and to clearly understand how different demographic groups rate each cuisine.

## 5. Filtering Demographic Data

Rows where all demographic information was missing were removed. This helps to focus the analysis on only those respondents who provided meaningful data.

# 6. Standardizing World Cuisines Interest Levels, and World Cuisines Knowledge Levels These variables were standardized into categories like "High," "Medium," "Low," and "None" to ensure consistency and for simpler segmentation of respondents.

#### 7. Adding Rating Categories

A new column called Rating Category was created based on the Rating column to simplify analysis. The ratings were categorized into "Low," "Medium," and "High," making it easier to group and interpret the data. Ratings of 1 and 2 were grouped as "Low," a rating of 3 as "Medium," and ratings of 4 and 5 as "High."

## 4.2 Data Analysis Techniques

Power BI Desktop was utilized to explore relationships between various variables, guiding the generation of meaningful insights through the calculation of key metrics. Additionally, the dashboard was designed to enable cross-filtering, allowing for interactive exploration of insights across the different demographic categories and cuisine preferences.

#### a. Key Metrics and Calculations

DAX (Data Analysis Expressions) was used to calculate key metrics and perform aggregations, which enabled dynamic calculations that adjusted based on the specific data points selected or the segments relevant within the dataset. The detailed DAX code and description of its purpose can be found in X. Appendix under the section Data Analysis Techniques.

Table 2. Measures Description

Measures	Description
Total Resp	The total number of unique respondents in the dataset.
%Female	The proportion of female respondents.
%Male	The proportion of male respondents.
Female	The number of female respondents.
Male	The number of male respondents.
Average Rating	The average rating of each cuisine.
Max Rating	Define the maximum rating score as 5.
Interest Level High	The number of rows that the "World Cuisines Interest Level" equals "High".
Interest Level Low	The number of rows that the "World Cuisines Interest Level" equals "Low".
Interest Level Medium	The number of rows that the "World Cuisines Interest Level" equals "Medium".
Interest Level None	The number of rows that the "World Cuisines Interest Level" equals "None".
Highly Rated Cuisine Proportion	The proportion of respondents rating each cuisine as "High."
Proportion of Respondents per HH Income	The proportion of respondents in each household income bracket.
Interest Level High	The number of rows that the "World Cuisines Interest Level" equals "High".

## 4.3 Dashboard Development

As outlined in the previous sections, the development of the dashboard started with importing and cleaning the data, organizing the dataset, and defining key relationships. We then created DAX measures to enable dynamic calculations. After which, a range of visualizations was used to present the analysis in a clear and accessible way. Common graphs, such as clustered bar charts for average cuisine ratings, stacked column charts for highly rated cuisines, pie charts for exploring respondents' knowledge and interest in world cuisines, and donut charts for rating distribution across demographics were created.

These visuals were designed to use cross-filtering, allowing users to interact with the data points and explore insights based on their selections. For example, filtering by gender or household income automatically updates charts, like the average cuisine ratings and world cuisine

knowledge levels. Additionally, interactive filters and slicers were also utilized, enabling users to filter the data and see how preferences and insights change across different U.S. regions.

## V. Dashboard Design and Visualisations

An exquisite visualization can raise audience interest and aid them to understand masterfully. We defined a visual aesthetic schema for the dashboard design. Referencing Microsoft Fluent Design guidelines, the functional and emotional design principle can be seen on the user interface all through (Fluent 2 Design System, n.d.). Different graphs and visual elements used in the dashboard were carefully selected to effectively communicate specific insights from the data.



Figure 1. Clustered Bar Chart for Average Cuisine Ratings

A clustered bar chart was used to provide a clear comparison of average ratings across different cuisines. It effectively shows how one cuisine compares to another in terms of respondent preference, allowing users to quickly identify which cuisines are most liked.

Interaction: Age and Gender and Highly Rated Cuisines charts, revealing how different demographic groups rate specific cuisines.



Figure 2.Stacked Column Chart for Highly Rated Cuisines

The stacked column chart displays the proportion of respondents who rated each cuisine as "High", focusing on cuisines where the proportion of high ratings is 60% or greater. Using this chart allows the users to quickly identify which cuisines are the most popular.

Interaction: Average Rating per Cuisine chart, to further highlight the popularity of highly rated cuisines across demographics.



Figure 3. Pie Chart for World Cuisines Knowledge Level

The pie chart was used to display the proportions of respondents based on their knowledge levels of world cuisines. Since there are only three categories represented as percentages, the pie chart is an effective choice that allows for a clear and simple visualization.

Interaction: Household Income and World Cuisine Interest Level to provide insights into which groups are more familiar with diverse cuisines.



Figure 4. Donut Chart for Rating Distribution

Donut charts were used to represent the distribution of ratings across different locations, filtered by specific cuisines. This visualization makes it easy to see which regions favor specific cuisines and how the ratings are distributed within each region.

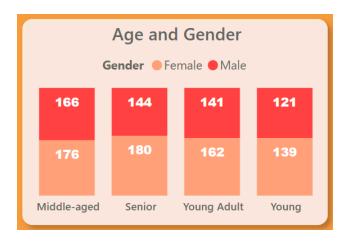


Figure 5. 100% Stacked Column Chart for Age and Gender Distribution

This 100% stacked column chart was chosen because it clearly presents the relative distribution of respondents by age group and gender, allowing users to quickly compare the demographic segments.

Interaction: Average Rating per Cuisine, Household Income, and World Cuisine Interest Level charts, to explore the preferences of different age and gender groups.



Figure 6. Clustered Column Chart for Household Income Distribution

The clustered column chart was chosen to simply and clearly display the proportion of respondents in each income bracket. It clearly shows the comparisons between the different income groups.

Interaction: Age and Gender, World Cuisine Interest Level, and World Cuisines Knowledge Level charts, providing insights into how income affects world cuisine interest and knowledge.

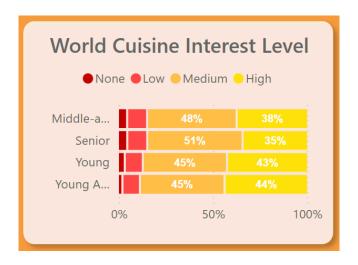


Figure 7.100% Stacked Bar Chart for World Cuisine Interest Level

This stacked bar chart was chosen to display the different levels of interest in world cuisines across different age groups. It is effective in presenting the proportional distribution of interests based on age at a glance.

Interaction: Household Income, Age and Gender, and World Cuisines Knowledge Level charts to examine the influence of age and income on interest in world cuisines.

#### 5.1 Visual Aesthetic and Creativity

According to Understanding Colour Psychology for Restaurants & Brands, orange is often associated with energizing, bold, and optimistic psychologically, and encourages sales in all sorts of dining areas, stimulating customers' appetite (Ashley Anastasia Howell, 2016). Therefore, we used consistent color schemes and labeling to enhance the user experience and ensure the dashboard was visually appealing. In detail, we adopt a highly saturated dark orange (#F69B3F) as the primary color, applying it to bars, the active state of components, and key information. A strong red (#FF4545) is the secondary color, which is used for less prominent components such as highlighted labels. A tomato red (#FFA73B) is the balance, playing the tertiary color role. With a vast font family, Segoe becomes the default typeface. Among them, the Segoe UI Semibold in the black color (#000000) is used for the title of each graph. All texts are right-aligned horizontally. Rectangle-shaped containers with 8 pixels corner radiuses are for the charts to give information prominence.

## 5.2 Dashboard Organization

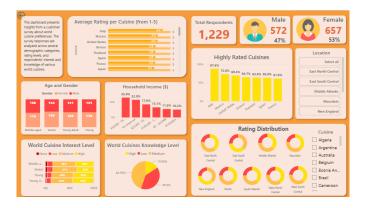


Figure 8. Dashboard

The layout was designed to ensure clarity and ease of use. Charts were logically grouped, with demographic information that can be used as filters placed on the left side. Key metrics, such as total respondents, gender distribution, and average ratings, were positioned at the top to provide a clear overview of the dashboard. More detailed insights into cuisine preferences, such as top rated cuisines and the distribution of ratings were positioned on the right side, allowing users to dive deeper into the data. Dynamic filters and slicers were also placed on the right, making it intuitive for users to interact with the dashboard and explore specific data points.

## VI. Findings and Insights

## 6.1 Descriptive Analytics

Descriptive analytics uses current data to identify trends and relationships, which simply describes trends and relationships (Cote, 2021). Although it doesn't dig deeper, it is useful for initial data analysis, playing a significant role as a springboard to drive further analysis and decision-making.

From customer profile statistics of respondents in the dashboard, respondents that belong to the middle-aged and senior age groups were more involved in the survey, and the female participation rate is slightly higher than male in all age groups. For the worldwide cuisine knowledge section, 44.75% of the respondents rated themselves as having a medium level of knowledge, while only 15.95% considered themselves highly knowledgeable. Surprisingly, in the high knowledge level segment, young adults are slightly more knowledgeable than middle-aged and senior respondents. Although the amount of younger respondents is mildly less in the sample, their interest in cuisines is significantly higher than that of the older group. People who live in the Pacific area have the highest participation rate and cuisine knowledge level. Among all cuisines in different countries, Italian cuisine, Mexican cuisine, and American cuisine receive the top 3 highest ratings. In particular, the proportion of high rated and average scores in Italian cuisine is remarkablely superior to others. Additionally, most of the respondents have household incomes ranging from \$50,000 to \$99,999, followed by those with an income between \$25,000 and \$49,999, ignoring those with the "no response" on this graph.

Preliminary analytics helped us to determine which regions and demographic groups show the highest interest in specific cuisines. All information extracted in this process can be used as a starting point for further analyses, providing insights into the "what happened" aspects of the dataset.

## 6.2 Diagnostic Analytics

Diagnostic analytics provides crucial insights about why a trend or relationship occurred, affecting strategy formulation directly. It can be viewed as a logical next step after using descriptive analytics to identify trends (Cote, 2021). This profound level of analysis answers the "why" aspects of the dataset, playing the most important role in decision-making and allowing businesses to pinpoint the factors contributing to their successes or challenges.

At the metrics World Cuisines Interest Percentage by Age Group, the proportion of respondents interest medium level has a slight upward tendency from 45% to 51%, while the high-level increases by 1% from the Young group to the Young Adult group, then decreases to 35% at the end. Through the clear differences in the distribution of this metric, we initially believe that there is a certain negative correlation between age and cuisine interest level, that is, the older the age, the lower the interest in cuisine.

At the metric Correlation between Household Income and Cuisine Interest (no response is ignored), in all interest categories, within the range of household income \$0 - \$99,999, the interest level goes up as the income increases. However, in the "High" and the "Low" interest categories, people's interest in worldwide cuisine goes down as income increases when their income is over \$99,999. Within the same income range, the interest declines and then rises. Therefore, we initially think that customer's interest in worldwide cuisine has a positive correlation with household income when their income is below \$100,000.

At the metric Correlation between Household Income and Cuisine Knowledge (no response is ignored), the number of all knowledge categories has the same tendency as the Cuisine Interest in the range of household income \$0 - \$99,999. The "Medium" and "Low" categories have the same downward trend when people's household income is beyond \$100,000, while the "High" category encounters a fluctuation. Similarly, we can initially consider that customers' knowledge of worldwide cuisine has a positive correlation with household income when their income is below \$100,000.

#### VII. Business Recommendations and Conclusion

To answer our scenario question, we recommend that the franchise expand its operations to the Pacific region of the United States due to its highest participation rate among all respondents in the survey, suggesting a strong potential for customer attraction. It is strongly recommended that Italian cuisine should be featured in the main menu for it has the best reputation around the world. Mexican and American cuisine can be offered as a secondary option. In addition, the menu can combine these 3 popular cuisines with local taste innovations to attract a wider customer base.

The franchise's market positioning aligns with public consumption levels, targeting consumers with household income between \$50,000 - \$99,999 from the middle-aged group aged 45-60. The data analysis shows that this demographic group has a relatively high interest and knowledge of world cuisines, and a huge consumption ability, which brings great potential business opportunities. Therefore, the marketing campaigns should focus on satisfying the preferences of this group, focusing on Italian cuisine while also offering Mexican and American dishes. While digital and social media campaigns can be used to target these groups and engage this particular audience, promoting loyalty programs or exclusive deals.

Younger demographics who show strong interest in global cuisines can also be engaged through interactive experiences like cooking classes or pop-up events. Combining these international cuisines with local flavors will attract and appeal to a wider audience.

#### VIII. References

Ledue, N. (2021, January 4). The impact of consumer preferences on food production. BOSS Magazine. https://thebossmagazine.com/changing-tastes/

Tastewise. (2024). Consumer buying behavior in the food industry. Tastewise.

https://tastewise.io/blog/consumer-buying-behavior-food-industry

Martínez-Ruiz, M. P., & Gómez-Cantó, C. M. (2016). Key external influences affecting consumers' decisions regarding food. Frontiers in Psychology, 7, 1618. https://doi.org/10.3389/fpsyg.2016.01618

Open Knowledge. (2020). Changing consumer preferences and food consumption patterns. Food and Agriculture Organization of the United Nations. https://openknowledge.fao.org

Amazon Web Services. (n.d.). What is ETL? ETL Explained - AWS. Amazon Web Services, Inc. https://aws.amazon.com/what-is/etl/

Fluent 2 Design System. (n.d.). Fluent 2 Design System. https://fluent2.microsoft.design/
Team Atlan. (2022, November 16). Data Dictionary: Examples, Templates, & Best practices. Atlan.com;
Atlan.

https://atlan.com/what-is-a-data-dictionary-to-test/#data-catalog-vs-data-dictionary-whats-the-difference

Ashley Anastasia Howell. (2016, July 15). Understanding Colour Psychology for Restaurants & Brands. Medium; Medium.

https://medium.com/@ashley\_howell/understanding-colour-psychology-for-restaurants-brands-dbb7ffbcec ae

Cote, C. (2021, November 9). What Is Descriptive Analytics? 5 Examples | HBS Online. Business Insights - Blog; Harvard Business School.

https://online.hbs.edu/blog/post/descriptive-analytics

Cote, C. (2021, November 18). What Is Diagnostic Analytics? 4 Examples | HBS Online. Business Insights - Blog. https://online.hbs.edu/blog/post/diagnostic-analytics

## Female icon in dashboard

popcornarts. (n.d.). Female icons. Flaticon. Retrieved from https://www.flaticon.com/free-icons/female

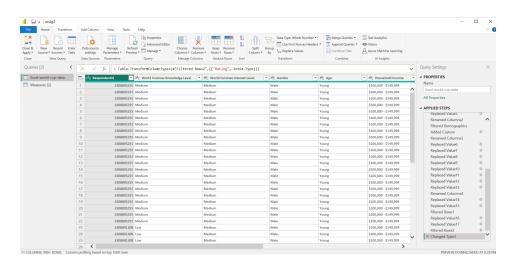
## Male icon in dashboard

popcornarts. (n.d.). Professional icons. Flaticon. Retrieved from https://www.flaticon.com/free-icons/professional

## IX. Appendix

#### 9.1 Dashboard Workflow

## a. Data Preprocessing



Sample image of the Power Query Editor where data cleaning and transformations took place.

## b. Data Analysis Techniques

#### 1. Total Respondents

```
Total Resp = DISTINCTCOUNT('food-world-cup-data'[RespondentID])
```

This value updates dynamically based on selections in the other charts, providing context by showing how many respondents are reflected in any filtered views.

## 2. Percentage of Female

```
%Female=DIVIDE(CALCULATE(DISTINCTCOUNT('food-world-cup-data'[Respo
ndentID]),'food-world-cup-data'[Gender]="Female"),CALCULATE(DISTIN
CTCOUNT('food-world-cup-data'[RespondentID]),ALL('food-world-cup-d
ata'[Gender]),0)
```

This value placed on cards dynamically changes based on the data point selected in any graph, allowing for real-time updates in the distribution of female respondents.

#### 3. Percentage of Male

```
%Male=DIVIDE(CALCULATE(DISTINCTCOUNT('food-world-cup-data'[Respond
entID]), 'food-world-cup-data'[Gender]="Male"), CALCULATE(DISTINCTCO
UNT('food-world-cup-data'[RespondentID]), ALL('food-world-cup-data'
[Gender]),0)
```

This value placed on cards dynamically changes based on the data point selected in any graph, allowing for real-time updates in the distribution of male respondents.

#### 4. Female

```
Female = CALCULATE([Total Resp],'food-world-cup-data'[Gender] =
"Female")
```

#### 5. Male

```
Male = CALCULATE([Total Resp],'food-world-cup-data'[Gender] =
"Male")
```

## 6. Average Rating

```
Average Rating = AVERAGE('food-world-cup-data'[Rating])
```

The purpose of this metric is to identify which cuisines were generally more preferred across the entire respondent pool.

#### 7. Max Rating

```
Max Rating = 5
```

#### 8. Interest Level High

The purpose of this metric is to gain insight on possible targeted marketing strategies and offerings tailored to specific age groups based on their interest levels.

#### 9. Interest Level Medium

#### 10. Interest Level None

## 11. Highly Rated Cuisine Proportion

```
Highly Rated Cuisine Proportion
=DIVIDE(CALCULATE(COUNTROWS(FILTER('food-world-cup-data','food-world-cup-data','food-world-cup-data')), Calculate(COUNTROWS('food-world-cup-data')), 0)
```

The purpose of this metric is to identify which cuisines have the highest proportion of favorable ratings (4 / 5), providing insights into which cuisines are the most popular among respondents.

## 12. Proportion of Respondents per HH Income

```
Proportion of Respondents per HH Income = DIVIDE([Total Resp],
CALCULATE(DISTINCTCOUNT('food-world-cup-data'[RespondentID]),
ALL('food-world-cup-data')),0)
```

#### 9.2 Dashboard Screenshot

