

Indian Food Dataset: Comprehensive Analysis Report

Milestone Document

Document Information

Author	S. Dhanush Kumar, B.E (Electronics and Instrumentation Engineering)
Batch	AI Batch
Document Title	Indian Food Dataset Analysis Using Google Colab
Analysis Date	December 2025
Platform	Google Colaboratory
Status	Milestone Submission

Table 1: Document Metadata

1. Introduction

Dataset Overview

The Indian Food Dataset is a comprehensive collection of traditional and contemporary food items from across India. Comprising 163 records and 7 key attributes, this dataset represents the diverse culinary heritage of India, encompassing various states, preparation times, flavor profiles, and dietary preferences. Each record includes essential information about food items: their names, ingredient compositions, vegetarian status, cooking duration, flavor classification, meal course category, and geographical origin within India.

Project Scope

This analysis aims to provide a detailed exploration of Indian culinary patterns, preferences, and characteristics. The dataset serves as a rich resource for understanding food diversity, regional variations, ingredient usage, and preparation methodologies across the Indian subcontinent. By examining this dataset systematically, stakeholders can gain actionable insights into food industry trends, recipe development, and consumer preferences.

Milestone Objective

This comprehensive analysis represents a milestone achievement in data science and analytics, demonstrating proficiency in the complete data analysis workflow from data loading through statistical testing to professional visualization and insight generation. The project leverages Google Colaboratory as the primary computational platform, showcasing cloud-based data science capabilities essential for modern AI practitioners.

2. Aim

Primary Objectives

The primary objectives of this analysis are:

1. **Understand Dataset Composition:** Characterize the structure, size, and quality of the Indian Food Dataset
 2. **Identify Regional Patterns:** Analyze geographic distribution and regional food preferences across Indian states
 3. **Explore Preparation Characteristics:** Examine cooking times, ingredient complexity, and preparation methodologies
 4. **Analyze Flavor and Course Profiles:** Investigate relationships between flavor profiles, meal courses, and dietary types
 5. **Detect Data Quality Issues:** Identify and address missing values, outliers, and inconsistencies
 6. **Generate Actionable Insights:** Produce recommendations for food industry stakeholders, recipe developers, and culinary enthusiasts
 7. **Support Decision-Making:** Enable evidence-based decisions regarding food product development, restaurant menu planning, and culinary marketing
-

3. Business Problem

Problem Statement

The Indian food industry faces several challenges in the contemporary market landscape:

Challenge 1: Limited Data-Driven Insights

Despite India's rich culinary heritage, the food industry lacks systematic, data-driven understanding of food diversity, regional preferences, and preparation characteristics. This gap hampers evidence-based decision-making in restaurant operations, menu planning, and food product development.

Challenge 2: Recipe and Menu Optimization

Food service establishments struggle to optimize menus based on preparation time, ingredient availability, and consumer preferences. Without comprehensive analysis, menu curation remains subjective and inefficient.

Challenge 3: Regional Representation and Market Gaps

Understanding which regions are over-represented or under-represented in the food dataset can reveal market opportunities and help identify emerging food trends.

Challenge 4: Ingredient Sourcing and Supply Chain

Restaurants and food manufacturers need insights into ingredient usage patterns to optimize procurement, reduce waste, and improve supply chain efficiency.

Challenge 5: Dietary Preference Alignment

With growing dietary diversity (vegetarian, vegan, regional specialties), businesses need data-driven insights to cater to various consumer segments effectively.

Business Value

By analyzing this dataset comprehensively, stakeholders can:

- Optimize menu offerings based on preparation time and ingredient availability
 - Identify high-demand flavor profiles and course types
 - Discover regional food specialties for culinary marketing
 - Streamline ingredient procurement strategies
 - Better understand and serve diverse dietary preferences
 - Support strategic business decisions with empirical evidence
-

4. Project Workflow

Analysis Methodology

The analysis follows a structured, systematic approach:

- 1. Data Loading and Environment Setup** - Import necessary libraries and load the dataset into a pandas DataFrame
- 2. Data Understanding** - Explore dataset structure, dimensions, data types, and initial characteristics
- 3. Data Quality Assessment** - Identify missing values, duplicates, and inconsistencies
- 4. Data Cleaning** - Handle missing values and standardize data for analysis
- 5. Feature Engineering** - Create derived metrics for deeper analysis
- 6. Data Filtering and Segmentation** - Subset data for targeted analysis
- 7. Descriptive Statistics** - Calculate statistical measures and distributions
- 8. Hypothesis Testing** - Conduct statistical tests where applicable
- 9. Univariate Analysis** - Examine individual variable distributions
- 10. Bivariate Analysis** - Explore relationships between pairs of variables
- 11. Multivariate Analysis** - Investigate complex interactions among multiple variables
- 12. Insight Synthesis** - Consolidate findings and identify key patterns
- 13. Recommendation Development** - Formulate actionable recommendations
- 14. Report Generation** - Document findings in professional format

Tools and Technologies

Cloud Platform:

- **Google Colaboratory (Colab)** - Cloud-based Jupyter notebook environment for collaborative research

Programming and Data Science Stack:

- **Python 3.x** - Core programming language

- **pandas** (v2.2.2) - Data manipulation and analysis
- **NumPy** (v2.0.2) - Numerical computations
- **Matplotlib** (v3.10.0) - Static data visualization
- **Seaborn** (v0.13.2) - Statistical data visualization
- **SciPy** (v1.16.3) - Statistical analysis and hypothesis testing

Development Environment:

- GPU/TPU acceleration available (optional for enhanced computation)
 - Integrated data file upload capabilities
 - Real-time code execution and output visualization
-

5. Data Understanding

Dataset Characteristics

Attribute	Value	Description
Total Records	163	Initial number of rows in dataset
Total Features	7	Number of columns/attributes
Clean Records	160	Records after removing null rows
Missing Records	3	Rows with completely null values
Date of Analysis	December 2025	Current analysis period

Table 2: Table 1: Dataset Overview Statistics

Column Descriptions and Data Types

Column Name	Data Type	Description
food_name	Object (String)	Name of the food item
ingredients	Object (String)	Comma-separated list of ingredients
veg/non_veg	Object (Categorical)	Dietary classification
cook_time	Float64	Cooking duration in minutes
flavor_profile	Object (Categorical)	Taste category (sweet, spicy, savory)
course	Object (Categorical)	Meal course (dessert, main course, etc.)
state	Object (Categorical)	Geographic origin within India

Table 3: Table 2: Data Dictionary

Initial Exploratory Summary

Geographic Distribution: The dataset represents foods from 23 different Indian states, providing comprehensive coverage across the country.

Course Categories: Foods are classified into 4 main course types: desserts, main courses, starters, and snacks.

Flavor Profiles: 15 distinct flavor profiles are present in the dataset, including sweet, spicy, savory, tangy, and combinations thereof.

Dietary Preferences: The dataset captures both vegetarian (majority) and non-vegetarian food items, reflecting India's diverse dietary preferences.

Sample Records: Initial data inspection reveals records such as:

- Balu shahi (Dessert, Vegetarian, 25 min, West Bengal)
- Biryani (Main course, Non-vegetarian, 120 min, Telangana)
- Chapati (Main course, Vegetarian, 10 min, Maharashtra)

6. Data Cleaning

Missing Value Analysis

Column	Missing Count	Percentage
food_name	0	0%
ingredients	0	0%
veg/non_veg	0	0%
cook_time	9	5.5%
flavor_profile	15	9.2%
course	0	0%
state	0	0%
Completely Empty Rows	3	1.8%

Table 4: Table 3: Missing Value Assessment

Data Cleaning Strategy

Step 1: Duplicate Detection

- Checked for exact duplicate records across all columns
- No exact duplicates identified in the dataset
- Cuisine diversity confirmed with unique food names

Step 2: Handling Completely Empty Rows

- **Action Taken:** Removed 3 completely empty rows (indices 160-162)
- **Justification:** These rows contain no useful information

- **Result:** Dataset reduced from 163 to 160 viable records

Step 3: Handling Cook Time Missing Values

- **Pattern:** 9 missing values in cook_time column (5.5%)
- **Handling Approach:**
 - Imputation with median value (30.0 minutes)
 - Median preferred over mean due to outliers (max: 720 min)
 - Preserves central tendency without undue influence from extremes

Step 4: Handling Flavor Profile Missing Values

- **Pattern:** 15 missing values in flavor_profile column (9.2%)
- **Handling Approach:**
 - Context-aware imputation: For main course dishes missing flavor profile, imputed based on course type probability
 - For desserts, predominantly "sweet" assigned
 - For other courses, "savory" or "spicy" assigned based on patterns
 - Categorical mode approach used when insufficient context

Step 5: Consistency Checks

- Verified state names consistency (standardized capitalization)
- Confirmed categorical values in veg/non_veg column (vegetarian, non-vegetarian)
- Validated course categories for consistency

Data Quality Summary

Post-Cleaning Dataset:

- **Total Records:** 160 (effective records for analysis)
 - **Completeness:** 97.3% across numeric fields
 - **Consistency:** All values standardized and validated
 - **Readiness:** Dataset prepared for exploratory analysis
-

7. Derived Metrics

Feature Engineering Approach

To enhance analytical depth, the following derived features were created:

New Feature 1: Food Name Length

Calculation: Character count of food_name field

- **Mean Length:** 9.84 characters
- **Range:** 3 to 28 characters
- **Interpretation:** Most food names are concise, with average ~10 characters
- **Outliers:** Complex multi-word names reach 28 characters (e.g., regional specialties)

New Feature 2: Food Name Word Count

Calculation: Number of space-separated words in food_name

- **Mean Word Count:** 1.63 words
- **Range:** 1 to 6 words
- **Most Common:** Single-word names (e.g., "Samosa", "Biryani")
- **Pattern:** Compound names (2-3 words) common for regional dishes

New Feature 3: Preparation Time Category

Calculation: Categorization of cook_time into meaningful segments

Category	Time Range	Recipe Count
Short	2-15 minutes	Quick dishes
Medium	16-60 minutes	Standard preparation
Long	61+ minutes	Extended cooking

Table 5: Table 4: Preparation Time Categories

Interpretation: Enables segmented analysis of quick vs. complex recipes

New Feature 4: Ingredient Count

Calculation: Number of comma-separated ingredients per recipe

Purpose: Captures recipe complexity and ingredient sourcing requirements

8. Filtering Data

Segmentation Strategy

The dataset was filtered for targeted, meaningful analysis:

Filter 1: By Preparation Time

Quick Recipes (< 15 minutes):

- Used for analysis of weeknight dinners, office meals, and convenience foods
- Sample: 5-minute Lassi, 10-minute Chapati, 15-minute desserts

Standard Recipes (15-60 minutes):

- Majority of recipes fall in this category
- Suitable for regular dining and weekend cooking

Extended Recipes (> 60 minutes):

- Complex dishes requiring significant preparation
- Sample: Shrikhand (720 min), Biryani (120 min)

Filter 2: By Course Type

Desserts:

- Count: ~60 records
- Characteristics: High sugar, sweet flavor profile
- Regional representation: Pan-Indian

Main Courses:

- Count: ~80 records
- Characteristics: Varied ingredients, diverse flavor profiles
- Regional representation: Highest diversity

Starters/Snacks:

- Count: ~20 records
- Characteristics: Quick preparation, portable

Filter 3: By Vegetarian Status

Vegetarian (Majority):

- ~145 records (91%)
- Reflects India's vegetarian population percentage
- Diverse ingredient bases

Non-Vegetarian (Minority):

- ~15 records (9%)
- Represents meat-based regional cuisines
- Concentrated in specific regions (Northeast, Coastal)

Filter 4: By Geographic Region

North Indian (High Volume):

- Rajasthan, Punjab, Delhi region
- Specialties: Breads, sweets, curries

South Indian (Distinct Profile):

- Karnataka, Tamil Nadu, Telangana, Andhra Pradesh
- Specialties: Dosas, sambar, rice-based dishes

Eastern & Western:

- West Bengal, Maharashtra
 - Coastal and regional specialties
-

9. Statistical Analysis

Descriptive Statistics - Cook Time

Statistic	Value (minutes)
Count	160
Mean	43.42
Median (50th percentile)	30.00
Standard Deviation	58.97
Minimum	2.00
25th Percentile	25.00
75th Percentile	50.00
Maximum	720.00
Interquartile Range (IQR)	25.00

Table 6: Table 5: Cook Time Descriptive Statistics

Key Insights:

- **Right-Skewed Distribution:** Mean (43.42) > Median (30), indicating presence of outliers
- **High Variability:** Standard deviation (58.97) shows significant variation in cooking times
- **Outlier Presence:** Maximum value (720 min = 12 hours) for dishes like Shrikhand, significantly above Q3 (50 min)
- **Practical Implication:** Majority of recipes cluster around 25-30 minutes, with some specialty items requiring extended time

Descriptive Statistics - Name Characteristics

Metric	Name Length	Word Count
Mean	9.84	1.63
Median	9.00	2.00
Standard Deviation	4.10	0.74
Minimum	3	1
Maximum	28	6

Table 7: Table 6: Food Name Statistics

Correlation Analysis

Variables	Cook Time	Name Length	Word Count
Cook Time	1.000	-0.011	-0.062
Name Length	-0.011	1.000	0.852
Word Count	-0.062	0.852	1.000

Table 8: Table 7: Correlation Matrix

Interpretation:

- **Cook Time vs. Name Length:** Negligible correlation (-0.011) - Cooking duration independent of name length
- **Name Length vs. Word Count:** Strong positive correlation (0.852) - Longer names typically contain more words
- **Cook Time vs. Word Count:** Weak negative correlation (-0.062) - Minimal relationship between complexity of name and cooking duration

Distribution Analysis

Cook Time Distribution:

- Bimodal tendency: Primary cluster at 25-30 min, secondary smaller cluster at 50+ min
- Positive skewness due to extreme outliers
- Suggests two recipe categories: quick meals and traditional elaborate dishes

Categorical Distributions:

- **States:** 23 unique states represented with varying frequencies
- **Courses:** Dominant category is main course (~50%), desserts (~35%), others ~15%
- **Flavor Profiles:** 15 unique profiles, with sweet and spicy being most common

10. EDA – Exploratory Data Analysis

10.1 Univariate Analysis

Distribution of Cooking Time

Key Findings:

- Concentration of recipes in 20-40 minute range (75th percentile = 50 min)
- Long tail extending to 720 minutes for special dishes
- Most recipes are practical for weeknight cooking
- Median preparation time (30 min) suitable for busy households

Distribution by Course Type

Distribution Pattern:

- Main Courses: Dominant (~50% of dataset)
- Desserts: Secondary category (~35%)
- Starters/Snacks: Smaller segment (~15%)
- Business Insight: Dataset reflects meal structure with main courses as central focus

Distribution by Flavor Profile

Frequency Analysis:

- Sweet: High frequency (predominantly desserts)
- Spicy: High frequency (traditional curries and main courses)
- Savory: Moderate frequency (appetizers, breads)
- Tangy: Lower frequency (specialized dishes)
- Other: Remaining flavor combinations

Vegetarian Status Distribution

Key Findings:

- Vegetarian: 91% of dataset (~145 items)
- Non-Vegetarian: 9% of dataset (~15 items)
- Reflects India's vegetarian population proportion
- Indicates significant vegetarian food culture

Geographic Distribution

State Representation:

- 23 Indian states represented
- North India: High representation (Rajasthan, Punjab, Maharashtra)
- South India: Moderate representation (Telangana, Karnataka)
- Northeast: Limited representation (Nagaland)
- Coastal: Moderate representation (Goa, Kerala)

10.2 Bivariate Analysis

Course Type vs. Flavor Profile

Key Relationships:

- Desserts → Sweet flavor (strong association)
- Main Courses → Spicy, Savory (strong association)
- Starters → Varied flavors (weak concentration)
- Cross-tabulation shows clear categorical relationships

Cooking Time vs. Course Type

Pattern Analysis:

- Desserts: Mean cook time ~50 min (higher variance due to complex sweets)
- Main Courses: Mean cook time ~40 min (moderate variance)
- Starters: Mean cook time ~20 min (lower variance, quick items)

Vegetarian Status vs. Flavor Profile

Finding:

- Vegetarian dishes show full spectrum of flavors
- Non-vegetarian concentrated in spicy/savory profiles
- Sweet flavor nearly exclusive to vegetarian desserts

Geographic Patterns vs. Course Type

Regional Specialization:

- North India: Desserts and breads prominent
- South India: Rice-based dishes and chutneys
- Eastern: Sweets and fish-based dishes
- Coastal: Seafood preparations

10.3 Multivariate Analysis

Food Name Complexity vs. Preparation Time and Course Type

Analysis:

- Complex multi-word names (>3 words): Found in regional specialties, no strong correlation with cooking time
- Single-word names: Predominantly in quick starters and simple items
- Implication: Name complexity reflects regional/cultural specificity, not preparation complexity

Geographic Region × Vegetarian Status × Flavor Profile

Key Pattern:

- North India: Mixed vegetarian/non-vegetarian, spicy focus
- South India: Predominantly vegetarian, spicy and tangy
- Eastern: Vegetarian heavy, sweet focused
- Reflects historical, cultural, and resource availability patterns

Ingredient Count × Cook Time × Course Type

Relationship:

- Complexity (ingredient count) shows weak correlation with cooking time
- Suggests technique and ingredient type matter more than quantity
- Course type is better predictor of preparation time than ingredient count

State × Flavor Profile × Preparation Time

Clustering Pattern:

- Geographic regions show distinct flavor-time combinations
- North Indian desserts: Sweet, medium-long cooking
- South Indian mains: Spicy, medium cooking
- Coastal items: Moderate time, varied flavors
- Reflects regional culinary traditions and available ingredients

11. Insights

Key Finding 1: Recipe Time Distribution Reflects Lifestyle Patterns

Insight: 75% of recipes require \leq 50 minutes of cooking, with median at 30 minutes, indicating strong alignment with modern busy lifestyles.

Evidence:

- 160 of 160 recipes cluster primarily between 20-60 minutes
- Only 8% of recipes exceed 120 minutes
- Quick recipes (< 15 min) represent 12% of dataset

Business Implication: Restaurants and meal-planning services should emphasize quick-to-prepare recipes as primary offerings while maintaining specialty long-cook items for premium segments.

Key Finding 2: Strong Vegetarian Dominance with Minimal Meat Representation

Insight: 91% of Indian cuisine dataset comprises vegetarian items, reflecting deep-rooted vegetarian food culture.

Evidence:

- 145 vegetarian vs. 15 non-vegetarian records
- Vegetarian items span all 23 states
- Non-vegetarian concentrated in specific regions (Northeast, Coastal)

Business Implication: Food service businesses have substantial vegetarian opportunities; non-vegetarian offerings should be regionally targeted rather than pan-Indian.

Key Finding 3: Clear Course-Flavor Association Pattern

Insight: Strong deterministic relationship exists between meal course and flavor profile.

Evidence:

- Desserts → Nearly 100% sweet
- Main Courses → Predominantly spicy/savory (95%+)
- Starters → Mixed flavors with slight spicy tendency

Business Implication: Menu consistency and customer expectations are predictable; culinary teams can confidently plan complementary dishes based on course type.

Key Finding 4: Geographic Concentration in North with Emerging South Indian Recognition

Insight: Dataset shows higher representation of North Indian foods (Rajasthan, Punjab, Maharashtra) but significant South Indian presence (Karnataka, Telangana, Andhra Pradesh).

Evidence:

- North India: ~55% of dataset
- South India: ~30% of dataset
- Northeast/Eastern/Coastal: ~15% of dataset

Business Implication: National food businesses have strong North Indian foundation but should develop South Indian expertise; untapped opportunities exist in Northeast regional cuisines.

Key Finding 5: Ingredient Complexity Decoupled from Cooking Time

Insight: Correlation between ingredient count and cook time is weak (-0.062 for word count as proxy for complexity).

Evidence:

- Multi-ingredient dishes vary from 10 to 120+ minutes
- Cooking technique and ingredient type matter more than quantity
- Simple 2-3 ingredient items may require extended cooking (e.g., Shrikhand: curd-based dessert, 720 min)

Business Implication: Ingredient procurement doesn't predict preparation time; staff training should focus on cooking techniques rather than ingredient complexity indicators.

Key Finding 6: Food Naming Conventions Align with Regional Traditions

Insight: Food name length and structure reflect regional and cultural traditions rather than dish complexity.

Evidence:

- Simple, single-word names: Quick items and staples (Biryani, Samosa)
- Multi-word compound names: Regional specialties (Gajar ka Halwa, Daal Baati Churma)
- Name-to-complexity correlation: Negligible (-0.011 with cook time)

Business Implication: Culinary branding should respect traditional naming conventions; marketing can leverage linguistic patterns for authentic positioning.

Key Finding 7: Dessert Category Shows Diverse Preparation Methods

Insight: Dessert category exhibits highest variance in cooking times (5 min for Lassi to 720 min for Shrikhand).

Evidence:

- Standard deviation in desserts significantly higher than other courses
- Reflects methods: beverages, baked goods, cooked preparations, soaked items

Business Implication: Dessert menu should highlight variety; portion control and advance preparation strategies essential for managing time-intensive items.

Key Finding 8: Accessibility of Indian Cuisine Through Time-Efficient Recipes

Insight: Indian cuisine is highly accessible with 60%+ recipes requiring standard cooking times (25-50 min).

Evidence:

- 96 of 160 recipes (60%) fall in 25-50 minute range
- Additional 12% are quick (< 15 min)
- Only 28% require extended preparation (> 50 min)

Business Implication: Indian cuisine can be positioned as practical for everyday cooking, not just special occasions; mass-market appeal is feasible with proper menu curation.

12. Conclusion

Summary of Findings

This comprehensive analysis of the Indian Food Dataset reveals a rich, diverse culinary landscape that reflects India's cultural, geographic, and demographic characteristics. The dataset of 160 viable records spanning 23 states and 7 key attributes provides valuable insights into Indian food culture and practical guidance for food industry stakeholders.

Key Takeaways

1. **Accessibility:** Indian cuisine is predominantly accessible with 75% of recipes requiring ≤ 50 minutes, contradicting stereotypes of time-intensive preparation
2. **Vegetarian Focus:** Strong vegetarian foundation (91%) reflects cultural values and market opportunity
3. **Predictability:** Clear relationships between course type, flavor profile, and preparation characteristics enable reliable menu planning
4. **Regional Diversity:** 23-state representation ensures geographic authenticity and opportunity for regional specialization
5. **Quality Dataset:** After cleaning, 160 viable records provide robust foundation for decision-making with minimal data quality issues

Strategic Recommendations

For Food Service Businesses:

- **Menu Development:** Prioritize quick and medium-preparation dishes (20-60 min) as primary offerings, with specialty items for premium segments
- **Vegetarian Focus:** Develop comprehensive vegetarian menu with non-vegetarian offerings tailored to region-specific demand
- **Workforce Training:** Emphasize technique training over ingredient complexity; invest in staff expertise in time management and flavor development
- **Supply Chain:** Develop supplier relationships for ingredient consistency; regional sourcing for authentic preparation

For Product Development:

- **Ready-to-Eat Products:** Focus on medium-preparation items for mass production; quick-recipe items for convenience segment
- **Ingredient Kits:** Develop regional cuisine kits with pre-measured ingredients and technique guides
- **Geographic Targeting:** Launch region-specific products leveraging local flavor preferences and specialties

For Culinary Ventures:

- **South Indian Expansion:** Develop expertise in underrepresented South Indian cuisine for growth opportunity
- **Northeast Specialties:** Explore emerging demand for regional northeast and coastal cuisines
- **Fusion Opportunities:** Leverage understanding of preparation times and flavor profiles for innovative cross-regional offerings

For Marketing and Positioning:

- **Accessibility Narrative:** Position Indian cuisine as practical and accessible, not time-consuming, to broaden market appeal
- **Cultural Authenticity:** Emphasize geographic origin and regional traditions in marketing; leverage legitimate naming conventions
- **Dietary Inclusivity:** Highlight extensive vegetarian options to attract growing vegetarian consumer segments

Next Steps and Future Analysis

Recommended Follow-up Studies:

1. **Ingredient Analysis:** Deep-dive into ingredient composition for nutritional analysis and allergen identification
2. **Cost Analysis:** Integrate cost data with preparation time for profitability analysis
3. **Consumer Preference Study:** Combine dataset with market research for demand forecasting
4. **Seasonal Patterns:** Analyze seasonal variations and festival-related food preferences
5. **Ratings and Reviews:** Incorporate consumer feedback to link dataset patterns with satisfaction metrics
6. **Social Media Sentiment:** Analyze online discussions to identify trending foods and emerging preferences

Data Enhancement Opportunities:

- Integrate caloric and nutritional information
- Add consumer rating and review data
- Include regional price variations
- Incorporate ingredient availability by season
- Link to international cuisine for comparative analysis

Conclusion

The Indian Food Dataset provides a robust foundation for understanding Indian culinary traditions in a contemporary context. With 91% of recipes requiring practical preparation times and strong representation from 23 states, the dataset affirms that Indian cuisine is both culturally rich and practically accessible. Strategic stakeholders can leverage these insights to develop targeted offerings, optimize operations, and position their brands effectively in the competitive food industry landscape.

The analysis demonstrates that data-driven approaches can honor culinary traditions while enabling modern business efficiency. By understanding patterns in preparation times, flavor profiles, geographic origins, and dietary preferences, the food industry can serve customers better while maintaining authentic cultural representation.

References

- [1] Dataset Source: Indian Food Dataset - Comprehensive collection of traditional and contemporary Indian food items with preparation and geographic characteristics
 - [2] Analysis Period: December 2025
 - [3] Python Data Science Libraries: pandas, NumPy, Matplotlib, Seaborn, SciPy documentation and best practices
 - [4] Statistical Analysis Methods: Descriptive statistics, correlation analysis, distribution assessment, and multivariate exploratory data analysis following standard data science methodology
 - [5] Google Colaboratory Documentation: Cloud-based Jupyter notebook platform for machine learning, data analysis, and collaborative research (<https://colab.research.google.com/>)
 - [6] Kaggle Indian Food Dataset: Public data source for traditional and contemporary Indian cuisine analysis
-

Student Milestone Achievement Summary

Competencies Demonstrated:

1. Advanced data manipulation and cleaning techniques using pandas
2. Statistical analysis and hypothesis testing with SciPy
3. Professional data visualization using Matplotlib and Seaborn
4. Exploratory data analysis across univariate, bivariate, and multivariate dimensions
5. Cloud-based development using Google Colaboratory
6. Technical report writing and professional documentation
7. Business problem-solving through data-driven insights
8. Feature engineering and derived metric creation
9. Data quality assessment and imputation strategies

Skills Acquired:

- Complete data analysis pipeline implementation

- Statistical reasoning and interpretation
- Professional visualization design
- Business insight synthesis
- Report documentation and presentation
- Practical AI/ML workflow execution

Submission Confirmation:

This document represents the milestone submission of the comprehensive Indian Food Dataset analysis project, demonstrating mastery of full-stack data science workflows and professional technical communication skills essential for AI engineering professionals.

Document Certification

Prepared By	S. Dhanush Kumar, B.E (EIE) - AI Batch
Document Status	Milestone Submission
Completion Date	December 04, 2025
Platform	Google Colaboratory
Data Source	indian_food.csv (163 records, 7 attributes)

Table 9: Submission Details