**Tasty Bytes**

***About the company***

2020:

* Founded in the midst of COVID Pandemic
* Started from inspiration and provide it
* Initially started as a recipe search engine
* Help people to find ways to use up limited supplies at home

2022+:

* Consolidates business
* Monthly subscription:
  + meet a full meal plan
  + The plan ensures a balance diet and meeting health too
  + Budget scalable for recipes
* Premium subscription:
  + All in monthly subscription
  + Deliver ingredients to your door

***Head Of Data Science Tasks***

**Project specifications**

* Create a written short report
* Summarize the analysis
* Thoughts of process of each step
* Code review
* Deliver a presentation
  + 8 - 10 slides
  + < 10 min
* Goal
  + Predict recipes which will be popular 80% of the time

**Written Report**

* Data Validation
  + Describe validation and cleaning for each column
* Exploratory data analysis for answering customer queries:
  + 2+ different graphics for single variables
  + 1+ graphics showing two or more variables
  + Describe findings
* Model Development
  + Type of problem
  + Baseline Model
  + Comparison Model
* Model Evaluation
  + Show how each model compares
* Definition of a metric for business to monitor
  + How should the business monitor what they want to achieve?
  + Estimate initial value(s) for metric based on current data?
* Final summary including recommendations that the business should undertake

**Presentation**

Present to the Product Manager (Product Owner)

* Project Overview
* Project Goals
* Work Summary to address the problem
* Key findings including metric to monitor
* Current Estimates
* Business Recommendaitons

**Product Manager**

* How the company did now
  + Choosing a favorite recibe and display on the page
  + Current max traffic has 40% of a popular receipt
* Assumptions
  + More traffic is more subscriptions
* Goal
  + Predict which recipes will lead high traffic
  + Correctly predict high traffic recipes 80% of the time
* Tasks
  + Present a solution soon

**Data Information**

* Data for each recipe and target variable (traffic)

|  |  |  |
| --- | --- | --- |
| Column Name | Type | Details |
| recipe | Numeric | Unique identifier of recipes |
| calories | Numeric | Number of calories |
| carbohydrate | Numeric | Amount of carbohydrates in grams |
| sugar | Numeric | Amount of sugar in grams |
| protein | Numeric | Amount of protein in grams |
| category | Character | Types of recipe. Recipes are listed on ten possible groupings  Lunch/Snacks  Beverages  Potato  Vegetable  Meat  Chicken  Pork  Dessert  Breakfast  One Dish Meal |
| servings | Numeric | Number of servings for the recipe |
| high\_traffic | Character | If the traffic to the site was high when this recipe was shown, this is marked with “High” |

* Revised

|  |  |  |
| --- | --- | --- |
| Column Name | Type | Details |
| recipe | Numeric | Unique identifier of recipes, no action |
| calories | Numeric | Number of calories, 52 null values that will be deleted |
| carbohydrate | Numeric | Amount of carbohydrates in grams, 52 null values that will be deleted. |
| sugar | Numeric | Amount of sugar in grams, 52 null values that will be deleted. |
| protein | Numeric | Amount of protein in grams, 52 null values that will be deleted. |
| category | Character | Types of recipe. Recipes are listed on eleven (11) possible groupings:   * Breakfast * Chicken Breast * Beverages * Lunch/Snacks * Potato * Pork * Vegetable * Dessert * Meat * Chicken * One Dish Meal |
| servings | Numeric | Number of servings for the recipe, no action |
| high\_traffic | Character | If the traffic to the site was high when this recipe was shown, this is marked with “High”. 373 null values that will be replaced with “Low”. |

**Grading**

|  |  |  |
| --- | --- | --- |
| **Competency** | **Sufficient** | **Insufficient** |
| **DATA VALIDATION** |  |  |
| Assess data quality and perform validation tasks | Has validated all variables and where necessary has performed cleaning tasks to result in analysis-ready data | Has not conducted all the required checks and/or has not cleaned the data. May have removed data rather than performed cleaning tasks |
| **DATA VISUALIZATION** |  |  |
| Create data visualizations in coding language to demonstrate the characteristics of data and represent relationships between features | Has created at least two different visualizations of single variables (e.g. histogram, bar chart, single boxplot)  Has created at least one visualization including two or more variables (e.g. scatterplot, filled bar chart, multiple boxplots)  Has used visualizations that support the findings being presented | Has used the same visualization throughout  Has not included graphics to represent single variables and relationships  Has not used visualizations that support the findings being presented |
| **MODEL FITTING** |  |  |
| Implement standard modeling approaches for supervised or unsupervised learning problems | Correctly identified the type of problem (regression, classification or clustering)  Has selected and fitted a model for that problem to be used as a baseline  Has selected and fitted a comparison model for the problem that they were provided | Has incorrectly identified the type of problem  Has not fitted a baseline model or has used a model for the wrong type of problem  Has not fitted a comparison model or has used a model for the wrong type of problem |
| **MODEL EVALUATION** |  |  |
| Use suitable methods to assess the performance of a model | Compared the performance of the two models/approaches using any method appropriate to the type of problem  Has described what the model comparison shows about the selected approaches | Has selected a method not suitable for the type of problem  Has not described what the results show about the selected approaches |
| **BUSINESS FOCUS** |  |  |
| Make recommendations for analytic approaches based on business goals | Has described at least one of the business goals of the project  Has explained how their work has addressed the business problem  Has provided at least one recommendation for future action to be taken based on the outcome of the work done | Has not identified any business goals  Has not explained how their work has addressed the business problem  Has not provided any recommendations for future actions |
| **BUSINESS METRICS** |  |  |
| Judge performance of analytic results against relevant business criteria | Has defined a KPI to compare model performance to business criteria in the problem  Has compared the performance of the two models/approaches using the defined KPI | Has not identified a KPI to compare the model performance to the business problem  Has not compared the performance of the two approaches using the defined KPI |
| **COMMUNICATION** |  |  |
| Employs multiple tactics (written and verbal) to communicate to business leaders | For each analysis step, has provided a written explanation of their findings and/or reasoning for selecting approaches  Has delivered a verbal presentation addressing the business goals, outcomes and recommendations | Has not provided a written summary for each step  Has not delivered a verbal presentation |