# Model Type + Complexity for Forecasting Collaborative Submissions

The following steps describe the process used to code participant submissions in terms of (1) type of forecast and (2) complexity of forecast.

## Model Type

Model Types are coded on a 1-4 scale, indicating the type of model the participants used for their prediction:

1. **Intuition** – These are forecasts where the participant made their predictions by relying solely on their intuition.
   1. Participants who make such forecasts will rely on their own expertise and their intuition to estimate how they expect the variable to change over the next 12 Months.
2. **Theory** – These are forecasts where the participant has indicated that their predictions were guided by a particular theory.
   1. Participants may indicate in their responses that their predictions were informed by a particular theory.
3. **Data-driven** – These are forecasts where the participant has indicated that they used one or more data-driven forecasting methods.
   1. Participants may indicate they used one or more methods of forecasting that can vary in complexity.
   2. These methods can be as simple as using the mean of the data provided, or as complex as an ARIMA or an econometric model.
4. **Mixed** – These are forecasts that used **Data-driven** methods **AND** either **Intuition** and/or **Theory.**
   1. If a participant indicates that they used both Theory and Intuition, but **NOT** Data, mark it as Theory.

**How to code** **Model Type**

Please follow these steps to determine how to code model type for a given row:

1. Look at the **Basis** column (Column D) to see how the participant categorized their own submission:
   1. The values in the basis column indicate the following:
      1. Intuition
      2. Theory
      3. Simulation
      4. Data-Driven
      5. Other
   2. At a minimum, the row can be coded similarly to what the participants inputted in this column
      1. ex. If a participant indicates in Column D that the basis for their forecast was Data-driven – a 4 – then this row can be coded as a **Data-Driven**, even if all other columns are empty.
   3. If the participant indicates Other (5) as a part of their basis, refer to Column E to see whether they provide any explanation.
2. Look at columns F (model), G (theory), & H (parameters) to determine whether the participant mentions the use of any Theory or Data-driven methods.
   1. Any mention of Mean is sufficient to code the row as **Data-Driven** (or **Mixed** if they also indicate **Intuition** and/or **Theory).**
3. If the columns are blank or lacking information, look at other rows that share a team name (column B) to determine whether any of the information in those rows are applicable to the current row.
   1. Ex. Some participants may have provided answers such as “See answers on Life Satisfaction”. Even though this particular row lacks information to code it, their row for life satisfaction indicates that they used mean and median alternatively (which would qualify as **Data-Driven**).
4. If you are unsure about how to code a particular row, mark the **Flag** column (column L) with a 1.
5. You may also write comments in column M to indicate your rationale or any questions about the row.

## Model Complexity

Model complexity is coded on a 1-3 scale, indicating the complexity of the forecasting methods used by participants to develop their forecasts:

1. **Simple forecasting method**
   1. Simple forecasting methods refers to forecasting methods that do not account for any additional parameters or variables, such as regression to the mean, Intuition-based forecasts, and/or simple Drift models.
2. **Moderate forecasting method**
   1. Moderate forecasting methods refer to forecasting methods that use 1-3 additional parameters to develop their forecasts, such as univariate time series forecasting models, Holt-Winters seasonal corrections, & auto-regressions with time lags.
3. **Complex forecasting method**
   1. Complex forecasting methods refer to forecasting methods that use more than 3 additional parameters to develop their forecasts, such as ARIMA and dynamic econometric models.

**How to code complexity**

Please follow these steps to determine how to code model complexity for a given row:

1. Look at the **Type** column (Column J) to filter between Data-Driven/Mixed and Intuition/Theory submissions.
   1. Any row that has a value of 1 or 2 for **Type** can automatically be given a value of 1 for **Complexity** (no need to look at the other columns in this case).
2. Look at columns F (model), G (theory), H (parameters), and I (number of parameters) to determine whether the participant mentions the type of forecasting method they used, and the number of parameters they used.
   1. **Please note**: the method used (ex. ARIMA) takes precedence in cases where a participant mentions a complex method, but does not list any (or few) additional parameters.
3. In some cases, participants may mention multiple types of models. If this occurs, try to identify the model they indicate they used for the actual forecast itself
4. If you are unsure about how to code a particular row, mark the **Flag** column (column L) with a 1.
5. You may also write comments in column M to indicate your rationale or any questions about the row.