**Proyecto 5: Modelo Predictivo de Tiempo de Vuelta - Protocolo Final Detallado**

**Objetivo:** Construir y validar un modelo de regresión para predecir el tiempo de vuelta (LapTimeSeconds) e identificar los factores más influyentes, utilizando JMP Standard.

**Fase 1: Preparación y Limpieza de Datos (El Setup del Coche)**

1. **Importar Datos:**
   * Abre JMPSummerChallengeDataset.csv en JMP, asegurándote de que se interprete el punto y coma (;) como separador y la coma (,) como decimal.
   * Guarda la tabla como un archivo JMP (.jmp) para preservar los tipos de datos y propiedades.
2. **Excluir Vueltas No Representativas:**
   * Abre el Filtro de Datos: Rows > Data Filter.
   * Añade filtros para las siguientes columnas y activa las casillas correspondientes para seleccionar las filas a excluir:
     + LapTimeSeconds -> Missing
     + PitInTimeSeconds -> Not Missing
     + PitOutTimeSeconds -> Not Missing
     + TrackStatus -> Selecciona todos los valores excepto 1.
     + Deleted -> Selecciona True.
     + LapNumber -> Selecciona 1.
   * Con las filas seleccionadas, ve a Rows > Hide and Exclude. Esto las eliminará de los análisis posteriores.
3. **Ingeniería de Características (Crear LapsOnStint):**
   * Crea una nueva columna: Cols > New Column....
   * Nómbrala **"LapsOnStint"**.
   * Haz clic en Column Properties y selecciona Formula.
   * Introduce la fórmula correcta y verificada:

Generated jsl

Col Rank( :Row(), :Driver, :Stint )

content\_copydownload

Use code [with caution](https://support.google.com/legal/answer/13505487" \t "_blank).Jsl

* + Haz clic en OK. Verifica que la columna se reinicia a 1 para cada nuevo stint de cada piloto.

**Fase 2: Validación y Modelado (La Simulación de Carrera)**

1. **Crear la Columna de Validación Manual:**
   * (Ya que no tienes Make Validation Column) Crea una nueva columna: Cols > New Column....
   * Nómbrala **"Validation"**.
   * Asígnale la siguiente fórmula para dividir los datos aleatoriamente (aproximadamente 75% para entrenamiento y 25% para validación):

Generated jsl

If( Random Uniform() < 0.75, "Training", "Validation" )

content\_copydownload

Use code [with caution](https://support.google.com/legal/answer/13505487" \t "_blank).Jsl

* + Haz clic derecho en la nueva columna Validation, ve a Column Info... y asegúrate de que el Data Type es Character y Modeling Type es Nominal.

1. **Construir el Modelo SOLO con el Conjunto de Entrenamiento:**
   * Abre una nueva tabla de datos solo con las filas de entrenamiento:
     + Usa el Data Filter para seleccionar las filas donde Validation es **"Training"**.
     + Con estas filas seleccionadas, ve a Tables > Subset.
     + Asegúrate de que está marcada la opción Selected Rows.
     + Nombra la nueva tabla **"Training Data"** y haz clic en OK.
   * **Trabaja ahora en la nueva tabla "Training Data".**
   * Lanza la plataforma de modelado: Analyze > Fit Model.
   * Configura los roles:
     + **Y:** LapTimeSeconds
     + **Effects (X):** Driver, Compound, Team, LapsOnStint, SpeedI1, SpeedI2, SpeedFL, SpeedST.
   * Haz clic en Run.
2. **Guardar la Fórmula del Modelo:**
   * En el informe de resultados que se genera, haz clic en el **triángulo rojo** superior (junto a "Response LapTimeSeconds").
   * Selecciona Save Columns > Prediction Formula.
   * Esto crea una nueva columna llamada Predicted LapTimeSeconds en tu tabla **"Training Data"**.

**Fase 3: Aplicación del Modelo y Medición de la Precisión (La Hora de la Verdad)**

1. **Preparar el Conjunto de Validación:**
   * Vuelve a tu tabla de datos **original y completa**.
   * Usa el Data Filter para seleccionar las filas donde Validation es **"Validation"**.
   * Crea un Subset de estas filas, nombrando la nueva tabla **"Validation Data"**.
2. **Aplicar la Fórmula al Conjunto de Validación:**
   * Ve a la tabla **"Training Data"**. Haz clic derecho en la cabecera de la columna Predicted LapTimeSeconds y selecciona Copy Column Properties.
   * Ve a la tabla **"Validation Data"**.
   * Crea una nueva columna vacía: Cols > New Column.... Nómbrala exactamente igual: Predicted LapTimeSeconds.
   * Haz clic derecho en la cabecera de esta nueva columna y selecciona Paste Column Properties. La fórmula se aplicará y se llenarán las predicciones.
3. **Calcular el RSquare de Validación:**
   * En la tabla **"Validation Data"**, abre Analyze > Fit Y by X.
   * **Y, Response:** LapTimeSeconds (el valor real).
   * **X, Factor:** Predicted LapTimeSeconds (el valor del modelo).
   * Haz clic en OK.
   * En el informe de resultados, busca la tabla Summary of Fit. El valor de **RSquare (ej. 0.7297)** es la precisión de tu modelo en datos no vistos. Anota este valor.

**Fase 4: Interpretación y Presentación (El Informe para Dirección)**

1. **Interpretar los Resultados del Modelo de Entrenamiento:**
   * Vuelve al informe de Fit Model generado a partir de la tabla **"Training Data"**.
   * **Effect Summary:** Identifica las variables estadísticamente significativas (p-value < 0.05). En nuestro caso: SpeedST, SpeedI1, SpeedI2, y Driver. Este es tu primer gran hallazgo.
   * **Profiler:** Abre el Profiler (Triángulo Rojo > Factor Profiling > Profiler). Prepara demostraciones interactivas para mostrar cómo cada variable (incluidas las no significativas como LapsOnStint y Compound) influye en el tiempo de vuelta predicho.
2. **Construir la Narrativa de la Presentación:**
   * **Introducción:** Presenta el objetivo de crear un modelo predictivo robusto.
   * **Metodología:** Explica tu riguroso proceso de limpieza de datos, la creación de LapsOnStint y, muy importante, el **proceso de validación manual** que realizaste para asegurar la fiabilidad del modelo.
   * **Resultados:**
     + **Validación:** Presenta tu **RSquare de validación (72.97%)** como la prueba de que el modelo es robusto. Muestra el gráfico de "Actual vs. Predicted" que generaste en la Fase 3.3.
     + **Factores Clave:** Muestra el gráfico Effect Summary y explica por qué las velocidades y el piloto son los predictores dominantes.
     + **Insights Ocultos:** Explica el concepto de colinealidad: por qué LapsOnStint y Compound no son significativos en el modelo de regresión, pero su efecto es real y visible en el Profiler.
   * **Demostración Interactiva:** Usa capturas de pantalla o un GIF del Profiler para mostrar visualmente el impacto de la degradación, los compuestos y el talento del piloto.
   * **Conclusión Estratégica:** Resume cómo este modelo validado, a pesar de sus matices, proporciona insights cuantitativos para la estrategia de carrera.