**Estrategia general del proyecto**

Te propongo una estructura para avanzar paso a paso:

**🔹 1. Análisis exploratorio (EDA)**

* Inspección general de las variables tabulares.
* Distribución de la variable objetivo (AdoptionSpeed).
* Chequeo de valores faltantes.

**🔹 2. Preprocesamiento**

* Codificación de variables categóricas.
* Tratamiento de texto (Description) y creación de embeddings o TF-IDF.
* Posible uso de metadatos e imágenes (más adelante si te interesa).
* Ingeniería de features (por ejemplo, total de colores, mezcla de razas, etc.).

**🔹 3. Modelado tabular**

* Empezar con modelos de clasificación (LightGBM, XGBoost, Random Forest).
* Evaluar performance (F1-macro, accuracy, Cohen's Kappa, etc.).
* Considerar métricas de orden si tratamos el problema como ordinal.

**🔹 4. Incorporación de texto**

* Procesar la columna Description con NLP.
* TF-IDF + modelos clásicos, o embeddings + redes neuronales.
* Incorporar análisis de sentimiento del archivo externo.

**🔹 5. (Opcional) Imágenes y metadata**

* Extraer features visuales o usar modelos preentrenados (ResNet, EfficientNet).
* Incorporar metadatos de Google Vision API.

**🔹 6. Stacking o ensemble**

* Combinar modelos tabulares, de texto e imagen para mejorar predicción.

**🔹 7. Entrega final**

* Preparar notebook limpio, bien comentado y con visualizaciones claras.
* Justificar decisiones, mostrar importancia de variables, y explicar resultados.