











<b>PREDICTION TASK</b>  <p>What is the prediction task?</p> <ul style="list-style-type: none"> <li>Tipo de tarea: Regresión supervisada.</li> <li>Entidad: Cada registro horario del sistema de bicicletas (día y hora).</li> <li>Resultado a predecir: Número total de bicicletas rentadas (cnt).</li> <li>Cuándo se observa el outcome: Al cierre de cada hora o día (según granularidad elegida).</li> </ul>	<b>DECISIONS</b>  <p>What are the decisions to be made?</p> <ul style="list-style-type: none"> <li>Cómo se usan las predicciones:</li> </ul> <p>Las predicciones alimentan un tablero de operaciones para planificar la reubicación de bicicletas y optimizar la disponibilidad por estación o zona.</p> <p>Si se predice baja demanda → reducir bicicletas disponibles.</p> <p>Si se predice alta demanda → enviar unidades extra o extender horarios.</p> <ul style="list-style-type: none"> <li>Parámetros del proceso: Umbrales de ocupación, clima, hora pico, día laboral.</li> </ul>	<b>VALUE PROPOSITION</b>  <p>What is the value proposition?</p> <ul style="list-style-type: none"> <li>Beneficiarios:</li> </ul> <ol style="list-style-type: none"> <li>Equipo de operaciones (mejor logística y menor costo de transporte).</li> <li>Usuarios finales (mayor disponibilidad y menos tiempos de espera).</li> </ol> <ul style="list-style-type: none"> <li>Pain points:</li> </ul> <ol style="list-style-type: none"> <li>Sobrecarga o escasez de bicicletas.</li> <li>Costos altos por reubicación manual.</li> </ol> <ul style="list-style-type: none"> <li>Integración / interfaz: Dashboard web con visualización de demanda esperada por hora o zona; alertas automáticas para operaciones.</li> </ul>	<b>DATA COLLECTION</b>  <p>How is the initial set of entities read?</p> <ul style="list-style-type: none"> <li>Origen inicial: Dataset público (UCI).</li> <li>Método: Descarga de CSV.</li> <li>Actualización: En un escenario real, mediante API del sistema de bicicletas o IoT de estaciones.</li> <li>Estrategia de frescura: Actualización diaria con pipeline automatizado.</li> <li>Costo: Bajo (datos abiertos).</li> </ul>	<b>DATA SOURCES</b>  <p>Where are the data sources?</p> <ul style="list-style-type: none"> <li>Internos: Base de datos de uso histórico del sistema (si se operara real).</li> <li>Externos: No Aplica</li> <li>Formato: CSV o tabular</li> </ul>
<b>IMPACT SIMULATION</b>  <p>What are the cost/benefit values?</p> <ol style="list-style-type: none"> <li>Costo/beneficio: <ul style="list-style-type: none"> <li>Falsa predicción de alta demanda → sobre costo de operación.</li> <li>Falsa predicción de baja demanda → pérdida de ingresos y satisfacción.</li> </ul> </li> <li>Simulación pre-despliegue: Backtesting sobre datos 2011–2012.</li> <li>Criterios de despliegue: <math>MAE \leq 45</math> y <math>RMSE \leq 70</math> en validación temporal.</li> <li>Fairness constraints: Evitar sesgos por temporada o clima extremo (balance de muestras).</li> </ol>	<b>MAKING PREDICTIONS</b>  <p>Are predictions made in batch or in real time?</p> <ul style="list-style-type: none"> <li>Modo: Batch (por día).</li> <li>Frecuencia: Cada día.</li> <li>Tiempo máximo: 5 min por batch diario.</li> <li>Recursos: dvc, kubeflow, bucket</li> </ul>		<b>BUILDING MODELS</b>  <p>How many models are needed?</p> <ul style="list-style-type: none"> <li>Número de modelos: 1 principal (regresión de demanda).</li> <li>Actualización: Semanal o mensual (según drift detectado).</li> <li>Tiempo de entrenamiento: &lt;30 min por pipeline.</li> <li>Recursos: Kubeflow, DVC, sklearn + LightGBM.</li> </ul>	<b>FEATURES</b>  <p>What representations are used for the data?</p> <ul style="list-style-type: none"> <li>Representaciones: Variables temporales, meteorológicas y de contexto.</li> <li>Transformaciones: <ul style="list-style-type: none"> <li>Normalización [0–1]</li> <li>One-hot encoding (season, weathersit, weekday)</li> <li>Features de rezago (cnt_lag_1h, cnt_lag_24h)</li> <li>Variables derivadas (is_weekend, is_peak_hour)</li> </ul> </li> </ul>
<b>MONITORING</b>  <p>What are the monitoring metrics?</p> <ul style="list-style-type: none"> <li>Métricas técnicas: RMSE, MAE, MAPE, drift (PSI/KS).</li> <li>KPIs de negocio:</li> <li>Reducción del 15–20% en redistribuciones ineficientes.</li> <li>Incremento de disponibilidad &gt;95%.</li> <li>Revisión: Semanal (métricas técnicas) y mensual (negocio).</li> <li>Herramientas: MLflow.</li> </ul>				