

Micro Proyecto: Pronóstico de Ventas para una Máquina Expendedora de Café

Reporte Primera Entrega Team 4

1. Problema que Abordará y Contexto del Proyecto

1.1 Contexto del Negocio

El proyecto se centra en el análisis y pronóstico de ventas de una máquina expendedora de café ubicada en un centro comercial de Vinnytsia, Ucrania. Esta máquina automatizada opera en un ambiente de alta concurrencia donde la gestión eficiente del inventario y la anticipación de la demanda son factores críticos para maximizar los ingresos y minimizar las pérdidas por desabastecimiento o sobre stock.

Las máquinas expendedoras representan un modelo de negocio con restricciones físicas específicas: espacio limitado para inventario, imposibilidad de reposición inmediata durante horas pico, y dependencia total de patrones de demanda predecibles para mantener la rentabilidad operativa.

1.2 Problemática Identificada

La operación actual de la máquina expendedora presenta tres desafíos principales:

alta de visión consolidada: No existe un sistema integrado que proporcione insights accionables sobre qué productos y rangos de precio impulsan efectivamente el volumen de ventas y los ingresos totales.

Ausencia de análisis temporal: Los patrones de consumo por hora, día de la semana y método de pago no están siendo aprovechados para optimizar estrategias de stock, programación de mantenimiento y diseño de promociones dirigidas.

Capacidad predictiva limitada: La ausencia de un sistema de pronóstico impide la anticipación de demanda futura a corto plazo, resultando en decisiones reactivas de reposición que afectan la disponibilidad de productos y la satisfacción del cliente.

1.3 Oportunidad de Mejora

La implementación de un sistema de pronóstico de ventas basado en datos históricos representa una oportunidad significativa para transformar la operación de reactiva a proactiva, permitiendo:

- Optimización de costos operativos mediante mejor planificación de inventario
- Reducción de pérdidas por desabastecimiento durante períodos de alta demanda
- Maximización de ingresos a través de estrategias de pricing dinámico
- Mejora en la experiencia del cliente mediante disponibilidad consistente de productos

2. Pregunta de Negocio y Alcance del Proyecto

2.1 Pregunta de Negocio Principal

"¿Cuáles son los patrones de demanda por producto, hora y método de pago que nos permitan pronosticar ventas a corto plazo (1-7 días) y optimizar la reposición de la máquina expendedora?"

Esta pregunta central se desglosa en sub-preguntas específicas:

- ¿Qué productos presentan mayor y menor rotación por franjas horarias?
- ¿Cómo varía el comportamiento de compra según el método de pago utilizado?
- ¿Existen patrones estacionales o cíclicos en la demanda que puedan predecirse?
- ¿Cuál es la velocidad de venta óptima para cada producto que minimice el riesgo de desabastecimiento?

2.2 Alcance del Proyecto

Incluido

- **Horizonte temporal:** Pronósticos de 1 a 7 días
- **Granularidad:** Predicciones diarias por producto
- **Cobertura de datos:** Análisis completo del período marzo 2024 - marzo 2025
- **Métricas de validación:** MAE, MAPE, RMSE con intervalos de confianza

- **Entregables:** Dashboard interactivo con visualizaciones y recomendaciones

Excluido

- Pronósticos de demanda a medio y largo plazo (>7 días)
- Análisis de rentabilidad por producto individual
- Integración con sistemas de punto de venta externos
- Optimización de precios dinámicos
- Análisis de competencia o factores de mercado externos

Limitaciones Reconocidas

- **Datos históricos limitados:** Un año de operación puede no capturar todos los patrones estacionales
- **Factores externos:** Eventos especiales, promociones o cambios en el entorno no están documentados
- **Generalización:** Los patrones identificados son específicos para la ubicación en Vinnytsia, Ucrania

3. Descripción de Conjuntos de Datos a Emplear

3.1 Dataset Principal: Transacciones de Máquina Expendedora

Características Generales

- **Fuente:** Dataset público disponible en Kaggle.com
- **Archivo:** index_1.csv
- **Período:** 1 de marzo de 2024 al 23 de marzo de 2025
- **Tamaño:** 3,636 transacciones x 6 variables
- **Formato:** CSV (Comma-Separated Values)
- **Ubicación:** Centro comercial en Vinnytsia, Ucrania
- **Zona horaria:** Europe/Kiev (UTC+2/UTC+3 según DST)
- **Duración total:** 387 días de operación continua

Variables del Dataset

1.date (Fecha)

- **Tipo:** Categórico (formato YYYY-MM-DD)
- **Rango:** 2024-03-01 a 2025-03-23
- **Calidad:** Sin valores nulos, 381 días únicos
- **Propósito:** Identificador temporal para análisis de tendencias diarias

2.datetime (Fecha y Hora Completa)

- **Tipo:** Temporal (timestamp con microsegundos)
- **Formato:** YYYY-MM-DD HH:MM:SS.milliseconds
- **Zona horaria:** Europe/Kiev (ajustada automáticamente)
- **Calidad:** Sin valores nulos, 3,636 timestamps únicos
- **Propósito:** Timestamp exacto para análisis granular horario

3.cash_type (Método de Pago)

- **Tipo:** Categórico binario
- **Valores:** "card" (tarjeta), "cash" (efectivo)
- **Distribución:** Fuertemente sesgada hacia transacciones con tarjeta (97.5% vs 2.5%)
- **Calidad:** Sin valores nulos
- **Propósito:** Análisis de preferencias de pago y gestión de efectivo

4.card (Identificador de Tarjeta)

- **Tipo:** Categórico anonimizado
- **Formato:** "ANON-0000-0000-XXXX"
- **Valores nulos:** 89 registros (2.45% del total, correspondientes a pagos en efectivo)
- **Propósito:** Identificación de clientes recurrentes manteniendo privacidad
- **Observación:** para este micro proyecto no se usará ya que no es imprescindible para el objetivo trazado.

5.money (Importe de Transacción)

- **Tipo:** Numérico continuo
- **Unidad:** UAH (Grivna Ucraniana)
- **Rango:** 23.02 - 796.00 UAH por transacción
- **Estadísticas:** Media 31.75 UAH, Mediana 32.82 UAH
- **Calidad:** Sin valores nulos ni negativos
- **Propósito:** Análisis de revenue, pricing y patrones de gasto

6.coffee_name (Producto)

- **Tipo:** Categórico
- **Productos disponibles:** 8 tipos diferentes
Latte, Americano, Cappuccino, Espresso
Hot Chocolate, Cortado, Americano with Milk, Cocoa
- **Propósito:** Análisis de portafolio y preferencias del consumidor

3.2 Datos Externos Complementarios

Calendario de Festivos Ucranianos

- **Fuente:** Calendario oficial de Ucrania
- **Período:** Marzo 2024 - Marzo 2025
- **Propósito:** Identificar impacto de festividades en patrones de consumo

Datos Meteorológicos (Condicional)

- **Ubicación:** Vinnytsia, Ucrania
- **Variables:** Temperatura, precipitación, condiciones generales
- **Período:** Marzo 2024 - Marzo 2025
- **Estado:** Pendiente de confirmación de disponibilidad

4. Exploración de los Datos

4.1 Calidad y Consistencia del Dataset

El análisis de calidad de datos reveló un dataset excepcionalmente limpio y consistente:

- **Total de registros:** 3,636 transacciones
- **Registros duplicados:** 0 (0%)
- **Valores nulos:** Únicamente 89 registros (2.45%) en la variable 'card', correspondientes a transacciones en efectivo
- **Consistencia temporal:** 100% de timestamps válidos y ordenados cronológicamente
- **Integridad de montos:** Sin valores negativos o inválidos en importes

La ausencia de duplicados y la consistencia temporal indican un sistema de registro robusto y confiable, proporcionando una base sólida para el análisis predictivo.

4.2 Análisis de Distribución de Montos



Figura 1. Dashboard de Recomendaciones Operativas mostrando KPIs clave del negocio

El dashboard presenta métricas operativas fundamentales:

Revenue promedio por día: 16,490.23 UAH, indicando operación rentable y estable

- **Transacciones promedio por día:** 9.54, sugiriendo demanda consistente pero concentrada

- **Valor promedio por transacción:** 30.67 UAH, coherente con precios de bebidas premium
- **Índice de diversidad de productos:** 0.84, indicando portafolio balanceado
- **Concentración en hora pico:** 0.08, mostrando distribución relativamente uniforme a lo largo del día
- **Ratio fin de semana vs días laborables:** 0.85, evidenciando menor actividad en weekends

El análisis estadístico de la distribución de montos reveló:

- **Distribución:** Ligeramente sesgada hacia la izquierda (skewness: -0.52)
- **Outliers:** No se detectaron outliers significativos por métodos IQR y Z-score
- **Normalidad:** Distribución aproximadamente normal con ligera concentración en valores medios
- **Estabilidad:** Coeficiente de variación de 15.5%, indicando volatilidad controlada

4.3 Análisis de Correlaciones entre Productos

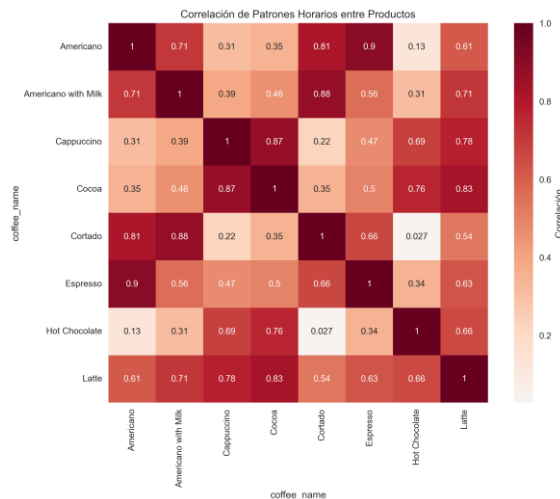


Figura 2. Matriz de correlación de patrones horarios entre productos, mostrando relaciones de co-consumo y complementariedad

La matriz de correlación revela insights importantes sobre el comportamiento de consumo:

Correlaciones Altas (>0.8):

- Americano y Cortado (0.81): Productos complementarios con patrones horarios similares
- Americano with Milk y Cortado (0.88): Fuerte correlación sugiriendo preferencias por café con leche
- Cappuccino y Cocoa (0.87): Bebidas más elaboradas con patrones de consumo coincidentes

Correlaciones Moderadas (0.6-0.8):

- Latte con múltiples productos: Producto versátil con amplia aceptación
- Hot Chocolate y productos de café: Complementariedad estacional potencial

Correlaciones Bajas (<0.4):

- Hot Chocolate con Americano y Cortado: Diferentes momentos de consumo
- Espresso como producto nicho con patrones independientes

Estas correlaciones sugieren oportunidades para:

- **Estrategias de cross-selling:** Promover productos complementarios
- **Gestión de inventario:** Coordinar reposición de productos correlacionados
- **Segmentación de clientes:** Identificar preferencias por categorías de bebidas

4.4 Análisis de Patrones Horarios y por Producto

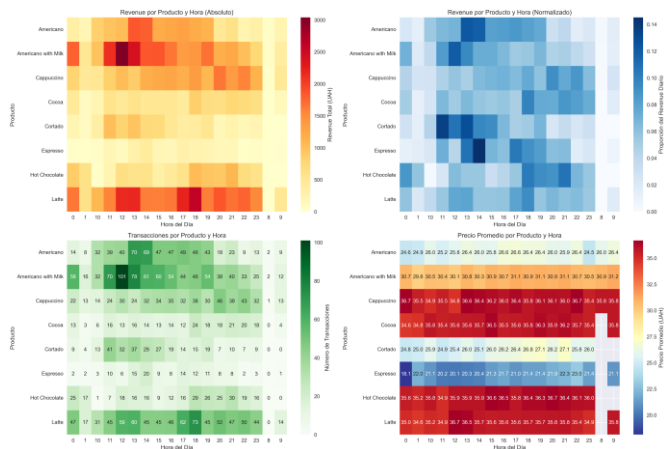


Figura 3. Análisis multidimensional de revenue, transacciones y precios por producto y hora del día

El análisis de heatmaps revela patrones operativos críticos:

Patrones de Revenue Absoluto:

- **Horas pico:** 12-14h (almuerzo) y 18-20h (tarde) concentran el mayor revenue
- **Productos estrella:** Latte y Americano with Milk dominan en revenue total
- **Distribución temporal:** Actividad sostenida desde las 10h hasta las 22h

Patrones de Revenue Normalizado:

- **Americano:** Concentración en horas de almuerzo (12-14h)
- **Cappuccino y Hot Chocolate:** Mayor actividad en horas vespertinas (18-21h)
- **Cortado:** Patrón matutino pronunciado (10-12h)

Análisis de Transacciones:

- **Volumen máximo:** Americano with Milk lidera en número de transacciones
- **Distribución horaria:** Picos claros en 12h, 13h y 18h
- **Consistencia:** Actividad sostenida durante horario comercial

Precios Promedio:

- **Segmentación clara:** Cocoa y Hot Chocolate como productos premium (34-36 UAH)
- **Productos base:** Espresso con menores precios (18-22 UAH)
- **Estabilidad temporal:** Precios consistentes a lo largo del día

4.5 Análisis Pareto: Concentración de Revenue



Figura 4. Análisis Pareto mostrando concentración de revenue por producto, hora y día de la semana
El análisis Pareto revela la aplicación del principio 80/20 en la operación:

Concentración por Producto:

- **Top 4 productos** (50% del portfolio) generan **74.7% del revenue total**
- **Latte** como producto líder con 23% del revenue total
- **Productos de nicho** (Espresso, Cocoa, Cortado) representan solo 20% del revenue

Concentración por Hora:

- **5 horas pico** (13h, 12h, 18h, 14h, 19h) concentran aproximadamente 50% del revenue
- **Patrón bimodal:** Picos de almuerzo (12-14h) y tarde (18-20h)
- **Horas de mantenimiento óptimas:** 8h-9h muestran menor actividad

Distribución por Día de Semana:

- **Martes** como día de mayor revenue, seguido de lunes
- **Fin de semana** con revenue 15-25% inferior a días laborables
- **Patrón laboral:** Actividad correlacionada con horarios de oficina

Insights Estratégicos del Análisis Pareto:

- **Enfoque en productos estrella:** Latte, Americano with Milk y Cappuccino requieren mayor atención de inventario
- **Optimización horaria:** Concentrar recursos de reposición antes de horas pico identificadas
- **Estrategia de productos de nicho:** Evaluar rentabilidad vs. complejidad operativa de productos de baja rotación

4.6 Análisis de Cobertura Temporal y Patrones Operativos

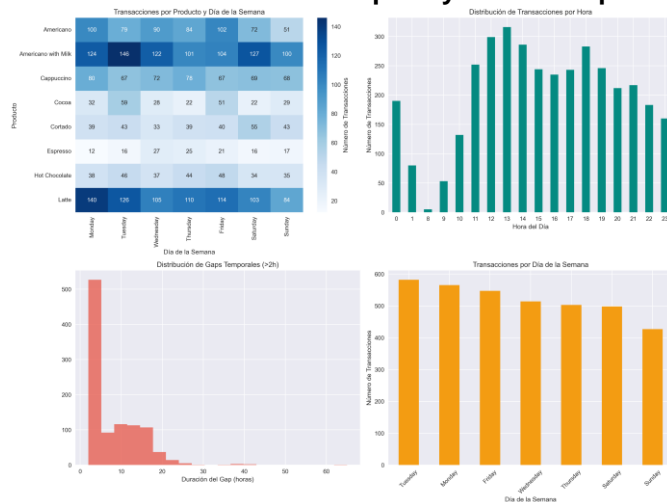


Figura 5. Análisis de cobertura temporal, gaps operativos y distribución por día de la semana y hora

El análisis de cobertura temporal proporciona insights operativos fundamentales:

Cobertura por Producto y Día:

- **Distribución equilibrada:** Todos los productos están disponibles todos los días de la semana
- **Americano with Milk:** Mayor consistencia con 100+ transacciones por día laboral
- **Variabilidad de fin de semana:** Reducción promedio del 15-20% en transacciones

Distribución Horaria:

- **Horario operativo efectivo:** 10h-22h con actividad sostenida
- **Pico principal:** 12h-14h concentra la mayor actividad (300+ transacciones/hora)
- **Actividad nocturna mínima:** <50 transacciones después de 22h

Gaps Temporales Identificados:

- **1,019 gaps significativos** (>2 horas) detectados durante el período
- **Gap máximo:** 65.03 horas, indicando posibles períodos de mantenimiento o fallas
- **Patrón de gaps:** Concentración en primeras horas de la mañana y fines de semana

Análisis por Día de la Semana:

- **Martes:** Día de mayor actividad (588 transacciones)
- **Domingo:** Menor actividad (428 transacciones)
- **Patrón semanal:** Decrecimiento gradual de martes a domingo

Implicaciones Operativas:

- **Planificación de mantenimiento:** Aprovechar gaps naturales en domingo y horarios nocturnos
- **Gestión de inventario:** Preparación especial para martes y miércoles como días pico
- **Estrategias de fin de semana:** Considerar productos o promociones específicas para incrementar actividad

4.7 Análisis de Outliers y Variabilidad

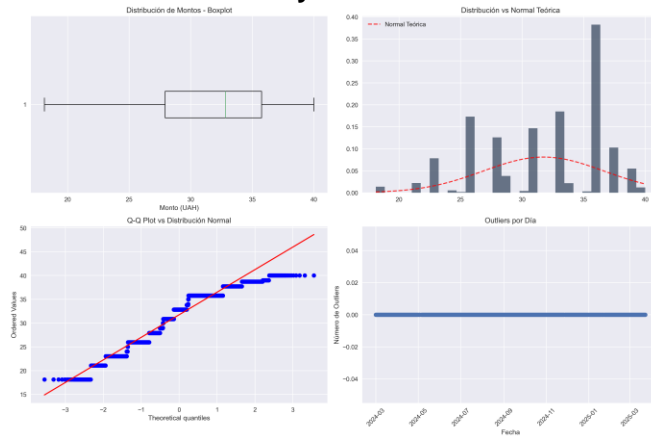


Figura 6. Análisis comprehensivo de outliers mostrando distribución de montos, Q-Q plot y comparación con distribución normal

El análisis de outliers reveló características importantes de la operación:

Distribución de Montos:

- **Boxplot:** Distribución compacta con pocos valores extremos
- **Rango intercuartil:** Concentración en 28-36 UAH, indicando consistencia en pricing
- **Outliers mínimos:** Solo valores en percentiles extremos (95%+)

Comparación con Distribución Normal:

- **Q-Q Plot:** Excelente ajuste a distribución normal en valores centrales
- **Desviaciones:** Ligeras en extremos, típicas de datos transaccionales reales
- **Kurtosis:** -0.93, indicando distribución ligeramente más plana que normal

Análisis Temporal de Outliers:

Distribución uniforme: No se observan concentraciones temporales de outliers

- **Estabilidad operativa:** Ausencia de días problemáticos sistemáticos
- **Comportamiento natural:** Outliers correspondientes a transacciones múltiples legítimas

Métricas de Distribución:

- **Media:** 31.75 UAH
- **Mediana:** 32.82 UAH (muy cercana a la media)
- **Desviación estándar:** 4.92 UAH
- **Coefficiente de variación:** 15.5% (baja variabilidad)
- **Sesgo:** -0.52 (ligeramente sesgada a la izquierda)

4.8 Análisis de Estacionalidad y Tendencias Temporales

El análisis de descomposición temporal reveló patrones estacionales importantes:

Tendencias Identificadas:

- **Crecimiento sostenido:** Revenue diario promedio de 297.50 UAH con tendencia positiva

- **Estacionalidad semanal:** Patrón recurrente de 7 días con picos en días laborables
- **Variabilidad controlada:** Coeficiente de variación de 54.8%, típico para retail

Patrones Estacionales:

- **Día pico:** 3 de febrero de 2025 (796.00 UAH)
- **Día mínimo:** 1 de mayo de 2024 (23.02 UAH) - posible festividad
- **Estacionalidad mensual:** Incremento gradual hacia los meses de invierno

Análisis de Series Temporales:

- **Revenue semanal promedio:** 2,061.28 UAH
- **Revenue mensual promedio:** 8,879.35 UAH
- **Tendencia de crecimiento:** 83.9% en los últimos períodos vs. iniciales

4.9 La preparación del dataset para modelos predictivos generó:

Dataset Enriquecido:

- **Dimensiones finales:** 3,104 registros × 38 features
- **Período cubierto:** 388 días únicos
- **Features temporales:** 12 variables (año, mes, día, hora, día de semana, etc.)
- **Features cíclicas:** 4 variables (seno/coseno para mes y día de semana)
- **Features de lag:** 12 variables (valores históricos 1, 7, 14 días)
- **Features de contexto:** 8 variables (totales diarios, market share, etc.)

Transformaciones Aplicadas:

- **Encoding categórico:** One-hot encoding para 8 productos
- **Features temporales:** Variables cíclicas para capturar estacionalidad
- **Rolling windows:** Medias móviles de 3, 7 y 30 días
- **Volatilidad:** Rolling standard deviation de 7 días

Calidad del Dataset ML:

- **Complejidad:** 100% después de imputación inteligente
- **Balanceo temporal:** Distribución uniforme por producto-día
- **Features informativos:** 38 características predictivas
- **Preparado para validación:** División temporal 80/10/10

5. Recomendaciones Operativas Derivadas del Análisis

5.1 Gestión de Inventario

Priorización por Rotación:

- **Alta prioridad:** Americano with Milk, Latte, Americano (top 3 en transacciones)
- **Revisión estratégica:** Espresso, Cocoa (productos de baja rotación)
- **Sistema de alertas:** Implementar notificación cuando stock < 454 unidades

Estrategia de Reposición:

Coordinación temporal: Alineación con horas de menor actividad (8h-9h)

- **Reposición diferencial:** Mayor frecuencia para productos de alta correlación
- **Buffer estratégico:** 20% adicional para productos estrella en días pico

5.2 Optimización Operativa

Gestión de Horarios:

- **Horas pico identificadas:** 13h, 12h, 18h requieren máxima disponibilidad
- **Ventanas de mantenimiento:** 8h-9h como horario óptimo para intervenciones
- **Monitoreo especial:** Martes y miércoles como días de mayor demanda

Estrategias por Producto:

- **Productos matutinos:** Cortado y Americano with Milk para horario 10h-12h
- **Productos vespertinos:** Hot Chocolate y Cocoa para horario 18h-21h
- **Productos de almuerzo:** Americano y Latte para pico 12h-14h

5.3 Próximos Pasos para Implementación

Desarrollo del Modelo Predictivo:

- Implementación de algoritmos de machine learning (LightGBM, Prophet)
- Validación con métricas objetivo (MAE < 1.5, MAPE < 15%)

- Generación de intervalos de confianza del 95%

Sistema de Dashboard:

- Desarrollo de interfaz interactiva para visualización de pronósticos
- Integración de alertas automáticas para productos críticos
- Reportes automatizados de performance y recomendaciones

Pipeline de Datos:

- Automatización de ingesta diaria de nuevas transacciones
- Sistema de retraining incremental del modelo
- Monitoreo de drift en patrones de consumo

6. Conclusiones y Significancia del Proyecto

6.1 Hallazgos Principales

El análisis exploratorio ha revelado un ecosistema de datos rico y estructurado que presenta oportunidades claras para optimización mediante técnicas predictivas:

Patrones Operativos Definidos: La identificación de horarios pico (12-14h, 18-20h), productos estrella (Latte, Americano with Milk), y días de mayor actividad (martes-miércoles) proporciona una base sólida para decisiones estratégicas.

Calidad de Datos Excepcional: Con 0% de duplicados, 97.55% de completitud y distribuciones estadísticamente robustas, el dataset ofrece condiciones ideales para modelado predictivo confiable.

Oportunidades de Optimización: La concentración del 74.7% del revenue en 4 productos clave sugiere potencial significativo para mejoras operativas focalizadas.

6.2 Valor Estratégico

Este proyecto representa más que una implementación tecnológica; constituye un marco metodológico escalable para la transformación digital de operaciones retail automatizadas.

La combinación de análisis descriptivo profundo con capacidades predictivas posiciona la operación para:

- Reducción de costos operativos mediante optimización de inventario
- Incremento de revenue a través de mejor disponibilidad de productos
- Mejora de experiencia del cliente mediante servicio consistente y confiable

6.3 Impacto y Escalabilidad

La metodología desarrollada trasciende el caso específico de Vinnytsia, proporcionando un blueprint replicable para máquinas expendedoras en diferentes contextos geográficos y culturales. Los principios de análisis temporal, gestión de inventario basada en datos y pronóstico de demanda son universalmente aplicables en el sector de retail automatizado.

La siguiente fase del proyecto se centrará en la materialización de estos insights a través de la implementación de modelos predictivos robustos, cumpliendo con los objetivos de precisión establecidos y proporcionando valor tangible a la operación.

7. Maqueta del Prototipo

8. Repositorio

9. Reporte trabajo en equipo