**PROYECTO ABP - ANÁLISIS DE DATOS**

**CARÁTULA**

**Institución:** Instituto Superior Politécnico de Córdoba (ISPC)

**Tecnicatura Superior:** Ciencia de Datos e Inteligencia Artificial

**Espacio Curricular / Módulo:** Exploración, Transformación y Limpieza de Datos utilizando Pandas

**Docente/s:** [Nombre del docente]

**Integrantes:**

* [Nombre Apellido 1]
* [Nombre Apellido 2]
* [Nombre Apellido 3]
* [Nombre Apellido 4]

**Ciclo Lectivo:** 2024

**1. NOMBRE | TÍTULO**

**"Análisis Estratégico de Evolución de Medios de Pago en Centros Comerciales de Estambul (2021-2023): Oportunidades de Captura de Market Share para Entidades Financieras"**

**2. TIPO**

**Proyecto Tecnológico de Investigación Aplicada**

**3. EJES | UNIDADES CONCEPTUALES**

**Unidad 1: Fundamentos de Programación y POO**

* Implementación de scripts Python para procesamiento de datos
* Uso de estructuras de datos (DataFrames, diccionarios, listas)
* Aplicación de conceptos de POO en manipulación de objetos Pandas

**Unidad 2: Fundamentos de Estadística y Probabilidad Aplicada**

* Medidas de tendencia central y dispersión (media, mediana, desviación estándar)
* Análisis de distribuciones de variables numéricas y categóricas
* Detección de valores atípicos en datos transaccionales

**Unidad 3: Análisis Exploratorio de Datos (EDA)**

* Identificación de patrones, tendencias y anomalías en series temporales
* Análisis univariado de medios de pago, tickets y frecuencias
* Análisis multivariado: correlación entre variables (edad, género, medio de pago, ticket)

**Unidad 4: Técnicas de Visualización de Datos**

* Histogramas para distribución de precios y cantidades
* Gráficos de línea para evolución temporal de medios de pago
* Mapas de calor para concentración temporal (día × centro × rubro)
* Boxplots para comparación de tickets por medio de pago

**Unidad 5: Base de Datos**

* Diseño de esquema relacional normalizado (clientes, ventas, productos)
* Consultas DML básicas (SELECT, WHERE, GROUP BY, ORDER BY)
* Consultas multitabla con JOIN
* Subconsultas para análisis de segmentos
* Procedimientos almacenados para cálculos recurrentes

**4. PROBLEMÁTICA | NECESIDAD | CASO**

Una entidad financiera regional con operaciones en Estambul enfrenta un ecosistema de pagos en transformación y alta competencia. La institución recopiló datos transaccionales de **99,457 operaciones** entre enero 2021 y diciembre 2023 en 5 centros comerciales estratégicos.

**Distribución General del Mercado (Promedio 2021-2023)**

* **Medios de Pago:** Efectivo 45%, Tarjeta de Crédito 35%, Otros medios digitales 20%
* **Demografía:** 60% Femenino, 40% Masculino, edades 18-65+ años
* **Rubros:** Ropa 35%, Cosméticos 15%, Otros 50% (5 categorías totales)
* **Características:** Tickets de $5 a $5,000, cantidades 1-5 unidades, 797 fechas únicas

**Desafíos Estratégicos**

1. **Evolución Dinámica:** Los promedios NO reflejan la realidad actual. Es necesario analizar cómo evolucionaron los comportamientos año tras año para detectar tendencias emergentes (crecimiento de billeteras digitales, descenso de efectivo).
2. **Competencia por Market Share:** El 35% que usa tarjeta está fragmentado entre múltiples emisores. La entidad necesita identificar dónde y cuándo concentrar descuentos para capturar clientes de la competencia.
3. **Conversión de Efectivo:** El 45% que paga en efectivo representa una oportunidad de bancarización. ¿Qué segmentos están migrando más rápido? ¿Dónde invertir en campañas de "primera compra con tarjeta"?
4. **Fidelización:** Maximizar el valor de clientes actuales identificando sus patrones de compra (rubros, centros, días preferidos) para diseñar programas de beneficios.

**Pregunta Central de Investigación**

¿Cómo evolucionaron los medios de pago entre 2021-2023 según segmentos demográficos, rubros y centros comerciales, y qué oportunidades estratégicas (captura de share vs. conversión de efectivo vs. adopción digital) deben priorizarse?

**5. FUNDAMENTACIÓN | HIPÓTESIS**

**Importancia del Problema**

El mercado de medios de pago en Turquía representa más de $150 mil millones USD anuales. La pandemia COVID-19 aceleró la adopción de pagos digitales, con crecimiento estimado de 300% en billeteras virtuales entre 2020-2023. Las entidades financieras que no detecten estas tendencias a tiempo pierden relevancia competitiva.

**Potencial del Proyecto**

* **Aplicación Práctica:** Genera recomendaciones accionables (calendario de campañas, segmentos prioritarios, alianzas comerciales)
* **Desarrollo Profesional:** Integra competencias técnicas (ETL, estadística, SQL) con pensamiento estratégico de negocios
* **Metodología Replicable:** El framework es aplicable a retail, e-commerce, telecomunicaciones y otros sectores

**Relevancia para el Perfil Profesional**

El proyecto desarrolla capacidades críticas del Técnico Superior en Ciencia de Datos e IA:

* Procesamiento de grandes volúmenes de datos (99,457 registros multidimensionales)
* Limpieza y preparación de datos reales con inconsistencias
* Análisis exploratorio para identificar patrones ocultos
* Visualización efectiva de insights complejos
* Traducción de hallazgos técnicos a lenguaje de negocio

**Impacto Social**

* **Inclusión Financiera:** Facilita acceso a servicios bancarios formales
* **Eficiencia Económica:** Reduce costos operativos y formaliza la economía
* **Transformación Digital:** Documenta adopción real de tecnologías financieras

**Hipótesis Principal**

El análisis temporal 2021-2023 revelará tendencias diferenciadas de evolución de medios de pago según segmentos, permitiendo identificar ventanas de oportunidad específicas con mayor probabilidad de éxito y ROI que estrategias genéricas no segmentadas.

**6. OBJETIVO GENERAL**

Desarrollar un análisis exploratorio integral mediante técnicas ETL para identificar y priorizar oportunidades estratégicas de captura de market share, conversión de clientes y adopción digital en el ecosistema de medios de pago de centros comerciales de Estambul, basándose en el análisis evolutivo (2021-2023) de 99,457 transacciones, generando recomendaciones accionables para optimizar inversiones en descuentos y alianzas comerciales.

**7. OBJETIVOS ESPECÍFICOS**

**OE1: Implementar Proceso ETL Completo**

Extraer, transformar y limpiar los datasets de ventas y clientes mediante Pandas, gestionando valores nulos, estandarizando formatos de fecha, validando tipos de datos, concatenando fuentes por id\_cliente, y cargando los datos procesados en una base de datos relacional SQL, documentando cada paso del proceso.

**OE2: Analizar Evolución Temporal de Medios de Pago y Segmentación**

Realizar análisis exploratorio de datos para identificar tendencias anuales/trimestrales (2021-2023) en la distribución de medios de pago, segmentando por demografía (género, edad), rubros de productos y centros comerciales, utilizando estadística descriptiva y técnicas de visualización para detectar patrones y anomalías.

**OE3: Identificar Concentración Temporal y Perfiles de Alto Valor**

Determinar días, horarios y temporadas de mayor concentración de transacciones por medio de pago, calculando tickets promedio y frecuencias de compra por segmento, y desarrollando un modelo de scoring que priorice oportunidades según tamaño de mercado, velocidad de cambio y valor transaccional.

**OE4: Generar Recomendaciones Estratégicas y Consultas SQL**

Traducir hallazgos en un plan estratégico con tres líneas de acción (captura de share, conversión de efectivo, adopción digital), replicar análisis clave mediante consultas SQL (multitabla, subconsultas, procedimientos almacenados), y elaborar informe ejecutivo con KPIs de seguimiento.

**8. CRONOGRAMA**

| **OBJETIVO ESPECÍFICO** | **ACCIONES, RECURSOS Y TIEMPOS** |
| --- | --- |
| **OE1: Implementar Proceso ETL Completo** | **Semanas 1-2**<br>• Descargar datasets de Kaggle (customer\_data.csv, sales\_data.csv)<br>• Carga en Pandas y análisis exploratorio inicial (info, describe, shape)<br>• Identificar valores nulos, duplicados, inconsistencias<br>• Limpieza: gestión de nulos, conversión de tipos, estandarización de fechas<br>• Concatenar dataframes por id\_cliente<br>• Diseñar esquema SQL normalizado y cargar datos limpios<br>**Recursos:** Python, Pandas, Jupyter Notebook, MariaDB<br>**Entregable:** Script ETL, DataFrame limpio, BD SQL poblada |
| **OE2: Analizar Evolución Temporal y Segmentación** | **Semanas 3-4**<br>• Crear variables temporales: año, trimestre, mes, día\_semana<br>• Análisis univariado: distribución de medios de pago por año<br>• Análisis multivariado: correlación entre variables<br>• Segmentación: método\_pago × género × edad × rubro × centro<br>• Calcular tasas de cambio (% variación 2021-2023)<br>• Visualizaciones: gráficos de línea, barras apiladas, heatmaps<br>**Recursos:** Pandas, Matplotlib/Seaborn, análisis estadístico<br>**Entregable:** Reportes de tendencias, matrices de segmentación, gráficos |
| **OE3: Identificar Concentración y Perfiles de Valor** | **Semanas 4-5**<br>• Análisis de concentración temporal (días, quincenas, estacionalidad)<br>• Calcular tickets promedio (precio × cantidad) por medio de pago<br>• Identificar top segmentos por volumen y valor<br>• Detectar valores atípicos (outliers) en precios<br>• Desarrollar scoring de oportunidades (tamaño + velocidad + valor)<br>• Priorizar top 10 oportunidades<br>**Recursos:** Pandas, boxplots, modelo de scoring<br>**Entregable:** Calendario de concentración, ranking de oportunidades |
| **OE4: Generar Recomendaciones y Consultas SQL** | **Semanas 6-7**<br>• Consolidar hallazgos de OE1-OE3<br>• Definir 3 estrategias con presupuesto sugerido<br>• Crear calendario de campañas y KPIs<br>• Replicar análisis con consultas SQL (JOIN, GROUP BY, subconsultas)<br>• Crear procedimientos almacenados para cálculos recurrentes<br>• Elaborar presentación ejecutiva e informe final<br>**Recursos:** SQL, síntesis de insights, comunicación<br>**Entregable:** Plan estratégico, scripts SQL, presentación, informe |

**DURACIÓN TOTAL:** 7 semanas

**9. PRODUCTO FINAL | CONCLUSIONES | RESULTADOS ESPERADOS**

**Entregables**

1. **Informe Técnico (10-15 páginas):** Proceso ETL documentado, análisis exploratorio con estadísticas descriptivas, visualizaciones, hallazgos por objetivo, conclusiones técnicas.
2. **Script de Código Comentado:** Notebook Jupyter con proceso completo, código limpio y documentado.
3. **DataFrames Procesados:** Archivo CSV limpio con datos unificados y diccionario de variables.
4. **Base de Datos SQL:** Esquema normalizado, scripts de creación, datos cargados, consultas que replican análisis (DML básico, multitabla, subconsultas, procedimientos almacenados).
5. **Presentación Ejecutiva:** 15-20 slides con resumen para stakeholders, visualizaciones de alto impacto, recomendaciones priorizadas.
6. **Plan Estratégico:** Documento con 3 estrategias (captura share, conversión efectivo, adopción digital), calendario de campañas, presupuesto sugerido, segmentos prioritarios, KPIs de seguimiento.

**Resultados Esperados**

**Cuantificación de Tendencias:**

* Tasa de crecimiento/descenso anual de cada medio de pago (2021-2023)
* Identificación de puntos de inflexión y aceleraciones
* Proyección simple de escenarios 2024-2025

**Segmentación Accionable:**

* 5-7 segmentos prioritarios claramente definidos (perfil, tamaño, potencial)
* Estrategia diferenciada por segmento demográfico y comercial

**Insights de Negocio:**

* "La Gran Migración": Cuantificar descenso de efectivo y hacia dónde migran
* "Amenaza Digital": Crecimiento de billeteras digitales en segmentos jóvenes
* "Zonas de Batalla": Días/horarios de concentración de transacciones con tarjeta
* "Diferenciación por Centro": Perfiles diferentes que justifican alianzas customizadas
* "Premio al Digital": Validar si tickets digitales son superiores a efectivo

**Recomendaciones Estratégicas:**

* **Estrategia 1 (40% presupuesto):** Captura de market share en días pico con descuentos agresivos
* **Estrategia 2 (35% presupuesto):** Conversión de efectivo mediante campaña "primera compra con tarjeta"
* **Estrategia 3 (25% presupuesto):** Integración digital (alianzas con billeteras virtuales, tarjeta virtual)

**10. BIBLIOGRAFÍA**

(Según Normas APA - completar con fuentes consultadas durante el proyecto)

* Documentación oficial de Pandas. https://pandas.pydata.org/docs/
* Dataset "Sales and Customer Data". Kaggle. https://www.kaggle.com/
* McKinney, W. (2017). *Python for Data Analysis*. O'Reilly Media.
* Wickham, H., & Grolemund, G. (2016). *R for Data Science*. O'Reilly Media.
* Date, C. J. (2019). *SQL and Relational Theory*. O'Reilly Media.
* [Agregar bibliografía adicional consultada sobre análisis de medios de pago, transformación digital financiera, etc.]