

Datathon 2025

Propuesta de Solución

Luis Daniel González Alcocer	A01286903
Hugo André Meza Fierros	A00841695
Alejandro José Alfaro García	A00842460
Gil Brandon García Contreras	A01254164

hey, banco



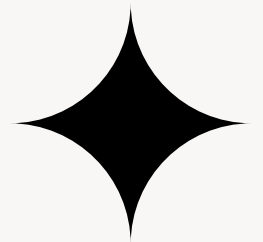
Índice

1. La Situación
2. Objetivos
3. Ideas Claves
4. Resultados
5. Limitaciones
6. Conclusiones





La Situación





¿Qué se necesita?

Un algoritmo y modelo para llegar a predecir las futuras compras de un usuario específico.

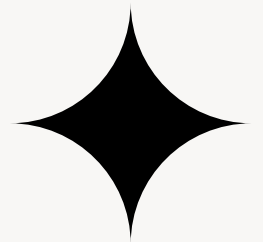
¿Cuándo es la siguiente compra de un cliente, su monto y el tipo de compra?



- Es posible ir más allá: gasto total por mes, número de transacciones, analizar frecuencia en establecimientos...
- Podemos hacer un análisis de regresión usando tecnología de punta y calcular la probabilidad de una posible compra para cada usuario
- Contamos con información de los hábitos de compra de 1 año de los clientes



Objetivos





Plan de Acción

Fase 1:

**Predecir los días
de los próximos
gastos.**

Fase 2:

**Predecir el tipo
de comercio y su
giro.**

Fase 3:

**Calcular el monto
esperado.**

**A partir de esto, se generará un reporte
para un usuario específico.**

*Esto permitiría a Hey Banco
ofrecer promociones personalizadas
justo antes de las compras recurrentes.*

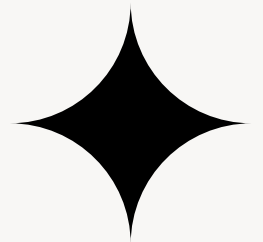
Permitiendo:

*Fidelizar al
cliente, que
percibe valor por
anticipación y
personalización.*

*Incrementar el
uso de la
tarjeta del
banco en
compras
planeadas.*



Ideas Claves





Se clasifican los tipos de gastos en tres diferentes tipos.

1. Gastos Recurrentes

Se sigue un patrón definido con intervalos de fechas y días promedio entre compras del mismo usuario.

Ejemplos: transporte, suscripciones, pagos de servicios, etc.

2. Patrones de gasto

Se consideran los patrones del pasado del usuario.

Ejemplo: Gastos personales, Comidas, Salidas a supermercados, Entretenimiento.

3. Predicción de futuros gastos

Se predicen gastos utilizando información de otros clientes para predecir gastos que el cliente todavía no efectúa.

Técnicas utilizadas:

Análisis de Patrones Recurrentes

- Identificación automática de gastos regulares por giro comercial
- Cálculo de periodicidad basado en historial transaccional
- Evaluación de regularidad y confiabilidad de patrones

Modelo Predictivo Híbrido

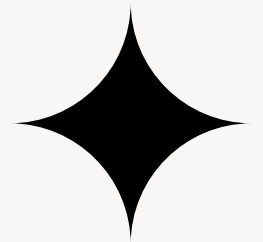
- Machine Learning, Análisis Estadístico, Filtros de Calidad

Métricas de Confianza

- Score Compuesto:** Frecuencia + Regularidad + Antigüedad del patrón
- Clasificación de Regularidad:** Alta/Media/Baja según variabilidad
- Validación Temporal:** Predicciones desde fecha actual hacia adelante



Resultados



Predicción Multidimensional:

- **Comercio específico** donde gastará el cliente
- **Fecha estimada** de la próxima transacción
- **Monto esperado** con base en historial
- **Probabilidad** de ocurrencia (score de confianza)
- **Periodicidad** de cada patrón identificado



Ejemplo de ejecución:

Analizando cliente: 8cf7b6a6cc74205009f63e37c438cc23181ee2c6

Periodo de predicción: 60 días

=====

INFORMACIÓN DEL CLIENTE:

- ID: 8cf7b6a6cc74205009f63e37c438cc23181ee2c6
- Tipo: Persona Física
- Actividad: JUBILADO
- Edad: 69 años
- Antigüedad: 6 años

RESUMEN DE PREDICCIONES:

- Gasto total esperado: \$1,524.83
- Comercios diferentes: 7
- Tipos de giro: 5
- Transacciones esperadas: 18

TOP 3 GASTOS ESPERADOS:

- COSTCO: \$176.18 el 2025-07-12 (Prob: 0.81)
- COSTCO: \$175.47 el 2025-05-28 (Prob: 0.81)
- COSTCO: \$165.54 el 2025-06-19 (Prob: 0.81)

Ejemplo de ejecución:



PREDICCIONES DETALLADAS:

comercio	fecha_predicha	dias_desde_hoy	monto_esperado	probabilidad	frecuencia_historica	regularidad
APLAZ	2025-05-26	1	123.70	0.448	37	Baja
OXXO	2025-05-26	1	35.98	0.434	48	Baja
ALLIANZ MEXICO	2025-05-27	2	269.30	0.765	29	Alta
7 ELEVEN	2025-05-27	2	37.55	0.599	48	Baja
APLAZ	2025-05-27	2	85.17	0.473	33	Baja
SMARTFIT	2025-05-27	2	66.41	0.445	11	Alta
KUESKI PAY	2025-05-27	2	181.82	0.099	12	Baja
UBER	2025-05-28	3	18.92	0.460	36	Media
DIDI RIDES	2025-05-30	5	13.48	0.901	68	Media
APLAZ	2025-05-30	5	130.92	0.448	37	Baja
AMAZON	2025-05-31	6	143.48	0.699	51	Media
DIDI RIDES	2025-05-31	6	66.25	0.389	18	Media
UBER	2025-06-01	7	11.16	0.460	36	Media
OXXO	2025-06-01	7	48.63	0.354	14	Media
APLAZ	2025-06-03	9	125.33	0.448	37	Baja
OXXO	2025-06-03	9	49.00	0.434	48	Baja
APLAZ	2025-06-04	10	106.90	0.473	33	Baja
DIDI RIDES	2025-06-05	11	15.92	0.901	68	Media
7 ELEVEN	2025-06-05	11	26.28	0.599	48	Baja
UBER	2025-06-06	12	8.80	0.460	36	Media
KUESKI PAY	2025-06-06	12	183.27	0.322	15	Baja
AMAZON	2025-06-07	13	40.73	0.699	51	Media
APLAZ	2025-06-07	13	123.59	0.448	37	Baja
DIDI RIDES	2025-06-07	13	6.41	0.389	18	Media
SMARTFIT	2025-06-09	15	64.92	0.445	11	Alta
DIDI RIDES	2025-06-10	16	14.62	0.901	68	Media
ALLIANZ MEXICO	2025-06-10	16	287.07	0.765	29	Alta

Ejemplo de ejecución:



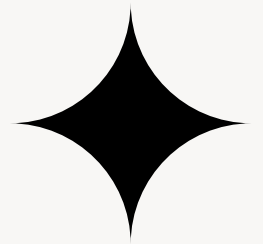
ESTADÍSTICAS:

- Gasto promedio por transacción: \$84.71
- Gasto total esperado: \$1524.83
- Confianza promedio: 0.80

=====



Viabilidad



Soluciones preexistentes:

Solución	Descripción	Diferenciador de esta Propuesta
Mint (Intuit)	Clasifica y predice gastos para consumidores con base en categorías amplias.	Este sistema va más allá, prediciendo <i>fecha específica</i> , <i>monto</i> y <i>comercio</i> individual por cliente.
Clearscore/Experian	Anticipan gastos recurrentes para ajustar líneas de crédito o evaluar riesgos.	La solución propuesta no solo anticipa riesgo, sino que habilita <i>acciones comerciales personalizadas</i> .
BBVA Smart Predictive Models	En algunos países, BBVA predice patrones de gasto. No disponible a cliente final con nivel de detalle individual.	Aquí se propone una solución <i>de uso interno</i> , con potencial de impacto en <i>segmentación</i> , <i>retención</i> y <i>marketing</i> .
Amazon Forecast	Predicción de series temporales multivariadas con infraestructura compleja (AWS).	La solución planteada es <i>liviana</i> , <i>interpretable</i> y <i>replicable</i> sin dependencia de infraestructura cloud.

Aproximación de costos:



Elemento	Costo Aproximado	Detalles
Desarrollo y validación (data scientist junior + senior – 3 semanas)	\$45,000 MXN	Basado en tarifas de mercado
Infraestructura (servidor local o nube básica)	\$0 – \$2,000 MXN/mes	Depende de escala (uso en pruebas o producción)
Licencias	\$0 MXN	Se usan solo librerías open source (Pandas, Sklearn)
Total estimado inicial	\$45,000 – \$50,000 MXN	

Costos Computacionales y Humanos:



Costo Computacional

Para un procesador AMD Ryzen 5 4600H se obtuvieron los siguientes datos:

Benchmarking para 100 ejecuciones del código:
206.14359169999807 segundos

Tiempo promedio por ejecución: 2.0614 segundos

Costo Humano:

- 1 científico de datos para mantenimiento/modelado.
- 1 analista para integración y supervisión periódica.
- 1 ingeniero para producción en entorno bancario (si se escala).



Limitaciones en el modelo:

Falta de integración con datos en tiempo real para análisis exacto.

No hay una incorporación de factores estacionales y económicos del país.

El modelo es independiente por cada usuario; dificultando su escalabilidad.

Se requieren compras de un usuario para llegar a predecir futuras compras.



Posibles mejoras:

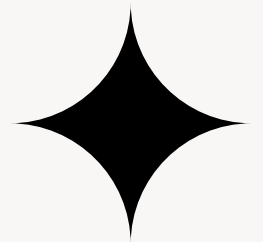
Explorar técnicas como LSTM, XGBoost o Prophet para capturar estacionalidad y patrones no lineales más complejos.

Entrenar modelos diferenciados según edad, perfil financiero o historial transaccional.

Tener alertas inteligentes al notificar al banco o al cliente cuando un gasto recurrente esperado no ocurre.

Considerar feriados, pagos quincenales, fechas de corte de tarjetas o promociones.

Conclusiones



Solución



El modelo efectivamente retorna una lista de compras esperadas en los futuros 30 días, prediciendo las fechas, el tipo de gasto, monto esperado y otras métricas.

PREDICCIONES DETALLADAS:

comercio	fecha_predicha	dias_desde_hoy	monto_esperado	probabilidad	frecuencia_historica	regularidad
APLAZ	2025-05-26	1	123.70	0.448	37	Baja
OXO	2025-05-26	1	35.98	0.434	48	Baja
ALLIANZ MEXICO	2025-05-27	2	269.30	0.765	29	Alta
7 ELEVEN	2025-05-27	2	37.55	0.599	48	Baja
APLAZ	2025-05-27	2	85.17	0.473	33	Baja
SMARTFIT	2025-05-27	2	66.41	0.445	11	Alta
KUESKI PAY	2025-05-27	2	181.82	0.099	12	Baja
UBER	2025-05-28	3	18.92	0.460	36	Media
DIDI RIDES	2025-05-30	5	13.48	0.901	68	Media
APLAZ	2025-05-30	5	130.92	0.448	37	Baja
AMAZON	2025-05-31	6	143.48	0.699	51	Media
DIDI RIDES	2025-05-31	6	66.25	0.389	18	Media
UBER	2025-06-01	7	11.16	0.460	36	Media
OXO	2025-06-01	7	48.63	0.354	14	Media
APLAZ	2025-06-03	9	125.33	0.448	37	Baja
OXO	2025-06-03	9	49.00	0.434	48	Baja
APLAZ	2025-06-04	10	106.90	0.473	33	Baja
DIDI RIDES	2025-06-05	11	15.92	0.901	68	Media
7 ELEVEN	2025-06-05	11	26.28	0.599	48	Baja
UBER	2025-06-06	12	8.80	0.460	36	Media
KUESKI PAY	2025-06-06	12	183.27	0.322	15	Baja
AMAZON	2025-06-07	13	40.73	0.699	51	Media
APLAZ	2025-06-07	13	123.59	0.448	37	Baja
DIDI RIDES	2025-06-07	13	6.41	0.389	18	Media
SMARTFIT	2025-06-09	15	64.92	0.445	11	Alta
DIDI RIDES	2025-06-10	16	14.62	0.901	68	Media
ALLIANZ MEXICO	2025-06-10	16	287.07	0.765	29	Alta

Elemento	Costo Aproximado
Desarrollo y validación (data scientist junior + senior – 3 semanas)	\$45,000 MXN
Infraestructura (servidor local o nube básica)	\$0 – \$2,000 MXN/mes
Licencias	\$0 MXN
Total estimado inicial	\$45,000 – \$50,000 MXN

Su enfoque híbrido combina la precisión del machine learning con la interpretabilidad del análisis estadístico, proporcionando predicciones accionables para estrategias de negocio.

