



Big Data For Bussiness

Proyecto fin de curso

Camilo Ferreira Fosalba
Faustino Harriague
Jorge Burgos De León
Jorge Caballero Bentancourt
Juan Manuel Bertón

- El **sector público** en Uruguay representa una gran parte de la inversión y el empleo. Tiene un alto grado de presencia en las actividades personales y económicas
- Antel: empresa monopólica
- Grado de madurez: Uruguay forma parte del Digital 9
- Líder regional del IDI por sexto año consecutivo
- Alta penetración de internet y telefonía móvil
- Iniciativas destacables: Ceibal, Ibirapitá



- AGESIC es la Agencia de Gobierno Electrónico y Sociedad de la Información y el Conocimiento
- Es el organismo encargado de trabajar en el desarrollo digital
- Depende directamente de la Presidencia y **trabaja de forma transversal** con las diferentes dependencias del gobierno
- Actualmente trabaja en desarrollo y gestión de **Portal Integrado de Gobierno (Gub.uy)** que se encuentra en marcha desde el 2018 y que recibe casi 1 millón de visitas mensuales



ID	Necesidad	Recurrencia	Viabilidad Técnica
1	Incentivar e incrementar visitas/uso portal	5	3
2	Mejorar intracomunicación entre entes estatales	3	2
3	Unificar criterios de diseño	2	5
4	Minimizar burocratización	2	2
5	Mejorar accesibilidad	4	4
6	Minimizar costos	1	3
7	Reposicionar imagen del estado	3	3
8	Mayor transparencia	2	2
9	Reducción o redistribución de personal	5	1
10	Mejorar experiencia de usuario	4	3

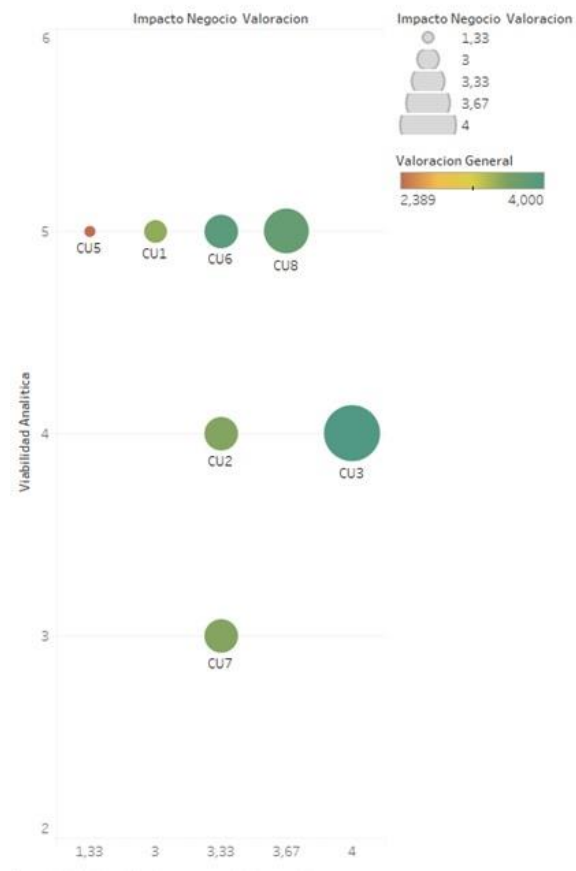
Casos de uso identificados

- CU1: Segmentación de usuarios para sugerencia de trámites y comunicados
- CU2: Análisis predictivo para detección picos de demanda de tramites
- **CU3: Segmentar perfiles de usuario en base a uso (revisando frecuencia de uso y que utilidad le dan al sitio) para fidelización de cliente**
- CU4: Identificación de anomalías en operativa
- CU5: Segmentación de organismos estatales más solicitados
- CU6: Segmentación de tramites más solicitados
- CU7: Evaluación de la atención presencial en servicios estatales y posibles sugerencias
- CU8: Evaluación de la plataforma del portal unificado y posibles sugerencias

Priorización de casos de Uso



ID	Monetización	Fuente de datos	Recurrencia	Viabilidad Técnica	Opt. operativa de la empresa	Incremento satisfacción cliente /usuario	Monetización nuevos productos / servicios
CU1	Interna	Interna	4	5	1	5	3
CU2	Mixta	Mixta	4	4	4	4	2
CU3	Interna	Interna	4	4	3	5	4
CU4	Interna	Externa	4	5	3	4	2
CU5	Interna	Interna	4	5	1	2	1
CU6	Mixta	Interna	5	5	2	4	4
CU7	Externa	Interna	5	3	5	3	2
CU8	Interna	Interna	3	5	4	5	2



CU3: Segmentar perfiles de usuario para fidelización



DESCRIPCIÓN

- Orientado a obtener una mayor fidelización de los usuarios e incrementar sus visitas al portal unificado

VALOR DE NEGOCIO

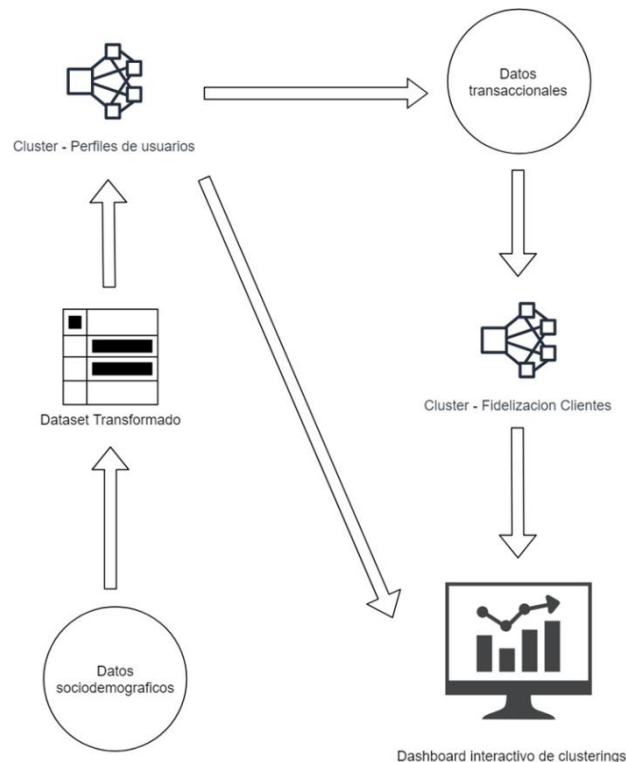
- Aumentar transacciones online (tramites/consultas)
- Disminuir atención personal a usuario (canal web y/o presencial en diferentes entes estatales)
- Aumentar eficiencias en comunicaciones, disponibilidad y ofertas de servicios
- Mejorar experiencia de usabilidad
- Aumentar satisfacción de usuarios
- Distribuir de manera eficiente los recursos de fuerza de trabajo a los componentes más críticos.
- Reducir fuerza de trabajo en base a las necesidades de mejoras necesarias identificadas

ENFOQUE ANALÍTICO: CLUSTERING

- GMM: Permite segmentar a los usuarios en función de su perfil sociodemográfico y de uso del portal. Por ejemplo la edad, educación, si tiene hijos, si es jubilado, y las variables relativas al uso
- K-Means: Se agrupan los datos que comparten patrones de tipos de usuario (extraído en base a análisis GMM), frecuencia de uso y tipo de uso. Aquellos que no pertenezcan a ningún cluster (outliers), serán usuarios no fidelizados.

CU3: Segmentar perfiles de usuario para fidelización

Diagrama analítico



Fuentes de Info

Comportamiento del cliente en el portal integrado (datos transaccionales/web)

- Cantidad de clics por sesión
- Elementos/áreas a las que ingresa
- Trámites iniciados y no iniciados por usuario
- Seguimiento de trámite por usuario
- Duración de tiempo de sesión
- Frecuencia de uso en (horas y días por semana)

Datos personales de clientes

- Edad
- Estado civil
- Dirección de residencia
- Lugar de nacimiento
- Ocupación
- Nivel de estudio

Infraestructura

Aprovisionamiento

WEB – SQL (base de datos)

Ingesta

Python - Sqoop

Almacenamiento

HDFS

Transformación/Procesamiento

Map Reduce - Hive SQL

Análisis

Python

Visualización

Tableau

- A partir del análisis de cluster obtendremos como resultado un número n de segmentos de usuarios. Esos segmentos deberían contar con perfiles determinados, combinando sexo / nivel educativo / edad / frecuencia de uso / servicios que consumen, etc.
- A partir de un dashboard que se alimente de dichos datos, **utilizaremos el análisis de cluster para casos particulares**, por ejemplo, **quienes son y qué perfil tienen aquellos que inician un trámite y no lo terminan**, o cuáles son las variables que explican el perfil de los usuarios frecuentes y que permitan identificar a los early adopters de un nuevo servicio del portal.
- **Combinado los clusteres con diferentes mapas, podemos analizar el comportamiento de diferentes regiones, como departamentos, municipios, etc.** Esto permitiría determinar los intereses y necesidades por zona geográfica.
- La información de estos mapas es útil para generar **políticas específicas** para zonas como por ejemplo talleres de capacitación para departamentos específicos, campañas de concientización, promociones,

Mockup del Dashboard



Roadmap



Acción	Descripción	Responsable	Semanas									
			1	2	3	4	5	6	7	8	9	10
Definición y ajuste	Definición y ajuste de las necesidades del negocio conjuntamente con el cliente	Business Consultant										
Kick off	Se definen roles y dedicación del equipo. Se acuerdan canales de comunicación	Business Consultant										
Relevamiento de datos	Relevamiento de fuentes de datos disponibles sobre los usuarios del portal	Business Consultant										
Diagnóstico de datos	Análisis de formato, calidad y actualización de esos datos	Data Architect										
Ingesta de los datos	Organización de los datos, limpieza e ingesta de los datos para su procesamiento	Data Ingeneer										
Procesamiento	Definir tipo de clusterización y hacer pruebas de clustering	Data Scientist										
Visualización	Diseño del visualizador/dashboard a ser utilizado e iteraciones con el científico de datos para validar visualización	Data visualization expert Data Scientist										
MVP Dashboard	Mínimo producto viable a presentar al cliente, para obtener feedback de la herramienta	Business Consultant Data visualization expert										
Producto Final (Dashboard)	Versión final del tablero/visualizador	Business Consultant Data visualization expert										
Sostenibilidad	Desarrollo junto con el usuario de los datos de una estrategia de uso del dato y apropiación	Business Consultant										

A definir