	<b>PRUEBA DE CONOCIMIENTOS DATOS NO ESTRUCTURADOS ViVi</b>		
	<b>VICEPRESIDENCIA BANCA PERSONAL Y MERCADEO</b> <b>ANALÍTICA - ViVi</b>		
	<b>Área: Analítica de Clientes</b>	<b>Elaborado</b> <b>15/09/2022</b>	<b>Versión:2</b>

## Prueba de Conocimiento ViVi

**Fecha de Inicio: Viernes 16 de Septiembre 2022, 6:00PM**

**Fecha Máxima de entrega: Viernes 23 de Septiembre 2022, 11:59PM**

**Entregar a:** [equipo.analitica.davivienda@gmail.com](mailto:equipo.analitica.davivienda@gmail.com)

En la organización buscamos personas con excelentes capacidades técnicas a las que les guste explorar los últimos avances en IA para asumir retos de especial dificultad. Aunque la siguiente prueba busca, en primera instancia, corroborar su idoneidad técnica, para nosotros es muy importante que esta destreza venga acompañada de dos elementos adicionales:

1. Creatividad para encarar desafíos técnicos en procesamiento de datos no estructurados.
2. Habilidad para comunicar el trabajo hecho y sus resultados a públicos generales y expertos.


Por lo tanto, el objetivo de la prueba es tener una visión completa de cómo es su compromiso, capacidad analítica, creatividad técnica y comunicación en la solución de un problema puntual de datos no estructurados, lo cual, en la práctica, se verá reflejado en su código y en una presentación a los jurados.

Para el desarrollo de esta prueba cuentan con una semana, tiempo en el que deben desarrollar la solución del problema mediante código debidamente documentado, depositado en un cuaderno de Python, así como una presentación gerencial de su trabajo en archivo pdf. De parte de los jurados se evaluarán estos dos elementos y se comprometen a respetar la propiedad intelectual de los desarrollos hechos por ustedes donde haya lugar (modelos personalizados, APIs propias, técnicas especiales, etc). Si lo desea, puede hacerlo explícito mediante licencia Creative Commons.

Puede hacer los supuestos que considere necesarios. Esto depende de la forma en que usted aborde el problema. Aunque en la literatura técnica se sugieren caminos, no hay una solución única para esta prueba. Fundamentalmente queremos ver desarrollos ordenados, creativos (si consideran que esta cualidad es pertinente en la resolución del problema) y bien comunicados.

### Objetivo

El propósito de la prueba es medir su capacidad de análisis, programación, elección de técnicas y comunicación de procesos y resultados. Dentro del Departamento de Datos No Estructurados creemos que la unión de estos elementos determina la buena calidad de las soluciones que le aportamos a la organización.

	<b>PRUEBA DE CONOCIMIENTOS DATOS NO ESTRUCTURADOS ViVi</b>		
	VICEPRESIDENCIA BANCA PERSONAL Y MERCADEO ANALÍTICA - ViVi		
	Área: Analítica de Clientes	Elaborado 15/09/2022	Versión:2

Nos interesa conocer de su propia voz el proceso que hizo para resolver el reto, ver la forma como abordó el problema, el orden en su código y la manera como sintetiza las partes interesantes de su trabajo en una presentación.


## El reto

Según Gartner, en este momento, los datos no estructurados conforman el 80-90% del total de los datos que manejan las compañías. Dentro de ellos, la información que está contenida en texto desestructurado (correos, decisiones judiciales, evaluaciones de clientes, comentarios en redes sociales, etc) representa un enorme potencial para desarrollar analítica dentro de las empresas.

Un objetivo evidente de procesamiento de texto no estructurado se encuentra en la analítica de opiniones. Mediante ella, podemos obtener información de primera mano de lo que piensan nuestros clientes, sin que dichas opiniones nos lleguen mediadas por encuestas u otros elementos logísticos que conformaban en un pasado reciente la única forma de tener información estructurada acerca de lo que hablan.

La meta de esta prueba es desarrollar, mediante herramientas de procesamiento de lenguaje natural (NLP), un análisis sobre un conjunto de tweets en los que se menciona de una u otra forma a nuestra organización. El análisis debe estar conformado por dos tareas principales:

1. **Una segmentación de los tweets** que permita conocer de qué más se opina en Twitter de nuestra organización, mediante la agrupación de los tweets con temáticas/dolores/opiniones comunes. Cada grupo de tweets encontrado debe tener una descripción, en lenguaje natural (i.e. no palabras clave, *wordclouds* o similares), que permita entender la temática central que los reúne en dicho grupo. Tenga en cuenta que, como esta tarea pretende emular un proceso analítico real que se llevaría a cabo periódicamente, tanto la segmentación como las descripciones para cada grupo las debe hacer el código hecho sobre cualquier conjunto de tweets que se le entregue como *input*. Así mismo, si se detecta que los datos pueden ser mejor explicados mediante una segmentación temática - subtemas, logrando de esta forma una mayor granularidad en la segmentación, ¡mejor aún!
2. **Una búsqueda de *insights*** en el conjunto de tweets suministrado, que puedan ser de interés para áreas como marketing, servicio al cliente, etc; y que sean visualizados de una forma clara para tomadores de decisiones en la presentación gerencial.


	<b>PRUEBA DE CONOCIMIENTOS DATOS NO ESTRUCTURADOS ViVi</b>		
	<b>VICEPRESIDENCIA BANCA PERSONAL Y MERCADEO</b> <b>ANALÍTICA - ViVi</b>		
	<b>Área: Analítica de Clientes</b>	<b>Elaborado</b> <b>15/09/2022</b>	<b>Versión:2</b>

Como se puede intuir, una cuidadosa ejecución de la tarea 1 puede hacer ganar mucho terreno en el desarrollo de la tarea 2. Sin embargo, recomendamos que la tarea 2 no esté cimentada exclusivamente en los resultados obtenidos en la tarea 1. Por ejemplo, el enfoque alternativo aportado por un buen EDA del conjunto de datos o la aplicación de múltiples técnicas NLP sobre los tweets pueden hacer que la cantidad y/o calidad de los insights obtenidos mejore ostensiblemente, al tiempo que pueden hacer destacar sus habilidades técnicas.

## Instrucciones y puntos importantes

- Se suministrará el conjunto de tweets a analizar.
- Los resultados y visualizaciones producidas por su desarrollo deben estar sustentados en la presentación.
- El desarrollo es libre, siempre y cuando quede contenido en el cuaderno entregado.
- Siéntase autónomo en usar las mejores técnicas que considere para resolver el reto. Puede utilizar cualquier herramienta, propia o que encuentre en Internet.
- El cuaderno debe tener las líneas de comando que instalen las librerías y versiones utilizadas para correr el código.
- El desarrollo hecho debe estar documentado, tanto en el código como en comentarios que nos permitan entender su elección de los pasos escogidos para la solución del análisis.
- Puede utilizar otros conjuntos de datos en su desarrollo, siempre que los resultados del análisis se limiten al conjunto de tweets proporcionados.
- En la presentación debe justificar las razones por las que escogió el enfoque desarrollado.
- Si ejecuta algún proceso previo o auxiliar en la solución del reto, por favor justifíquelo en la presentación.
- Si por algún motivo su desarrollo no tiene los resultados esperados por usted, recuerde que esto también es un resultado. Elabore la presentación y justifique su proceso. ¡Es posible que su trabajo sea **muy interesante y tenga mucho potencial!**

**Bono:** Adicionalmente, si tiene conocimientos de despliegue, “apificación” o puesta en producción de su desarrollo, puede hacerlos explícitos en la presentación mediante un bosquejo sustentado de lo que usted haría para que su desarrollo sea utilizado por un usuario final. Esto reemplazará el peor puntaje parcial que haya obtenido en la calificación de su presentación. El despliegue propuesto no se tiene que desarrollar.

	<b>PRUEBA DE CONOCIMIENTOS DATOS NO ESTRUCTURADOS ViVi</b>		
	VICEPRESIDENCIA BANCA PERSONAL Y MERCADEO ANALÍTICA - ViVi		
	Área: Analítica de Clientes	Elaborado 15/09/2022	Versión:2

## Insumos

Para el desarrollo se compartirá un correo con el asunto

**‘Prueba\_de\_conocimientos\_cientifico\_davivienda\_ViVi’**

en el cual se encontrará adjunto el archivo zip **prueba\_NLP\_cientifico\_vivi.zip** con el siguiente contenido:

1. Archivo plano **davivienda\_tweets.csv** con el conjunto de tweets a analizar.
2. El presente documento explicativo.

## Entregables


Se debe entregar a través de correo un archivo zip, con el formato de nombre:

**vivi\_cientifico\_datos\_no\_estructurados\_nlp\_[nombre del Participante].zip**

el cual contenga una carpeta con los siguientes elementos:

- **Presentación Gerencial:** en formato *pdf*, donde se encuentre:
  - Exploración de datos.
  - Técnicas usadas o exploradas.
  - Dificultades encontradas.
  - Aspectos importantes.
  - Resultados.
  - Costos de la ejecución de su desarrollo, si corresponde (Ej: si utilizó una API paga, una VM con hardware especial, etc).
  - Detalles adicionales que considere relevantes para la presentación de su proceso.
  - (Opcional) - Bosquejo de puesta en producción del desarrollo hecho.
- **Códigos:** Cuaderno de Python suministrado con el desarrollo hecho, debidamente documentado. Los resultados de la ejecución de cada celda deben ser visibles sin necesidad de correr las celdas. Tenga en cuenta que no vamos a modificar rutas de ubicación de input/output, archivos auxiliares, etc. por lo que todo el contenido debe ser completo y consistente dentro del cuaderno entregado.

**NOTA: Puede hacer su entrega completa a través de un repositorio de Github. En dado caso, por favor notificar por correo el vínculo y verificar que el repositorio compartido sea público para su evaluación.**

	<b>PRUEBA DE CONOCIMIENTOS DATOS NO ESTRUCTURADOS ViVi</b>		
	<b>VICEPRESIDENCIA BANCA PERSONAL Y MERCADEO</b> <b>ANALÍTICA - ViVi</b>		
	<b>Área: Analítica de Clientes</b>	<b>Elaborado</b> <b>15/09/2022</b>	<b>Versión:2</b>

- **(Opcional)** - Si su desarrollo necesita de archivos auxiliares para funcionar, por favor inclúyalos dentro del repositorio de manera que sea consistente con la ejecución de los códigos hechos.

Los resultados deben entregarse por medio de correo electrónico dentro del plazo estipulado al principio de este documento a [equipo.analitica.davivienda@gmail.com](mailto:equipo.analitica.davivienda@gmail.com)

## Evaluación

Se evaluarán sus capacidades de análisis, manejo de datos, programación e interpretación. Esto se realizará bajo la siguiente metodología:

- Desarrollo de los códigos: **50%**
  - Orden (20%)
  - Documentación (10%)
  - Técnica (20%)
- Presentación con los resultados: **50%**
  - Orden en el relato del proceso (15%)
  - Conocimientos técnicos (10%)
  - Completitud en los datos relevantes de su desarrollo (10%)
  - Claridad en la presentación (15%)
- **Bono:** Si demuestra suficiencia de conocimientos de despliegue de la analítica presentada, obtendrá un buen puntaje en el bono y éste reemplazará la peor calificación que tenga en los siete criterios anteriores.