

***Proyecto Bases de Seguros***

***Banco del Norte S.A.***

***Procesamiento de Datos***

***Luis Miguel Barragán Merchán***

***Febrero de 2022***

Contenido

[1. Identificación del Problema. 3](#_Toc94375909)

[1.1 Identificación de un problema. 3](#_Toc94375910)

[1.2 Investigación. 3](#_Toc94375911)

[1.3 Búsqueda de Soluciones Anteriores. 3](#_Toc94375912)

[2. Planteamiento de Preguntas. 4](#_Toc94375913)

[2.1 El Problema. 4](#_Toc94375914)

[2.2 Preguntas. 4](#_Toc94375915)

[3. Colección de Datos. 4](#_Toc94375916)

[3.1 Fuente. 4](#_Toc94375917)

[4.Analisis Exploratorio del Dataset. 4](#_Toc94375918)

[4.1 Cargar DatFrame. 4](#_Toc94375919)

[5. Limpieza de Datos y Agregaciones. 5](#_Toc94375920)

[5.1 Limpiar la Data de NaN. 5](#_Toc94375921)

[5.2 Generar Agregaciones. 5](#_Toc94375922)

[6. Automatización. 5](#_Toc94375923)

[6.1 Automatización con foor loops. 5](#_Toc94375924)

[6.2 Manejo de errores con try except. 5](#_Toc94375925)

# 1. Identificación del Problema.

## 1.1 Identificación de un problema.

Dentro del área de seguros de un banco pequeño latinoamericano que hace muy pocos años ha iniciado sus operaciones, se encuentra el proceso de carga de bases de datos para los asesores comerciales telefónicos a través de un archivo de formato Excel, dicha base contiene los datos de los clientes que tienen vehículos con prenda a favor del cliente y que tienen por vencer la póliza de seguro de estos vehículos.

Esta base se genera mediante un proceso de extracción de información a partir de otros archivos de formato Excel histórico que se han ido guardando a través del tiempo de operación. Se realiza un proceso puramente manual que acarrea bastante tiempo de procesamiento de un colaborador, además se está expuesto a errores por una falla de quien ejecuta el procesamiento.

Por otra parte, se quiere automatizar el procesamiento de la información por medio de un procesamiento de datos en Python a través de Jupyter-Lab, así como cargar la información en tablas dentro de una base de datos SQL Server para ir depurando los archivos históricos de Excel y realizar el proceso de manera automática.

## 1.2 Investigación.

Dentro de las operaciones del banco muy pocos procesos se encuentran automatizados, dado que antes de convertirse en un banco eran una compañía de financiamiento que manejaba casi toda la mayoría de sus procesos en archivos planos por la cantidad de clientes que se tenían, actualmente la cantidad de clientes ha ido en aumento; dado lo anterior se requiere con suma urgencia la automatización de varios procesos, entre ellos las bases de seguros.

## 1.3 Búsqueda de Soluciones Anteriores.

Según antecedentes de la compañía se encuentran desarrollando estos procesos mediante macros en archivos de formato Excel habilitados para macros, como una posible forma de automatizar dichos procesos. Adicionalmente se planea migrar algunos, no todos los procesos a un software a la medida desarrollado en lenguaje Java.

# 2. Planteamiento de Preguntas.

## 2.1 El Problema.

¿Como automatizar el procesamiento de un archivo Excel, cargar su información en tablas de una base de datos relacional para posterior consulta de datos y estadísticas a través de un dashboard apoyados en lenguaje Python?

## 2.2 Preguntas.

* ¿Qué cantidad de clientes se tienen actualmente con créditos de vehículo en la entidad bancaria?
* ¿Cuáles son las aseguradoras con las cuales se encuentran asegurados los clientes y que porcentaje de clientes maneja cada una?
* ¿Cuáles son los concesionarios con mayores colocaciones de créditos de vehículo?
* ¿Cuáles son las marcas de vehículos mas solicitados por los clientes?
* ¿Cuáles son los servicios (particular, público, carga, etc) con mayor crédito de vehículo en la compañía?

# 3. Colección de Datos.

## 3.1 Fuente.

La fuente de los datos con la cual se trabajara es un archivo Excel con la siguiente nomenclatura:

* Base\_Seguros\_Banco\_Comercial.xlsx

Se cargaran los datos a una base de datos relacional desarrollada en SQL Server con el siguiente nombre:

* Seguros\_BdN

# 4.Analisis Exploratorio del Dataset.

## 4.1 Cargar DataFrame.

# 5. Limpieza de Datos y Agregaciones.

## 5.1 Limpiar la Data de NaN.

## 5.2 Generar Agregaciones.

# 6. Automatización.

## 6.1 Automatización con foor loops.

## 6.2 Manejo de errores con try except.