# EL CONTEXTO

Uno de los puntos más importantes de tu investigación es saber explicar cómo de importante es tu estudio. Seguramente tienes muy claro que tu investigación tiene mucho sentido y puede aportar mucho valor.

Pero no es suficiente. Es muy importante saber transmitir a tu audiencia lo importante e interesante que es tu estudio.

Para ello, te he preparado 3 preguntas que te van a ayudar a expresarlo. Además te van a servir para definir mucho mejor el objetivo de tu investigación.

Trata de contestar a estas preguntas con palabras muy sencillas que un niño de 12 años pueda entender.

¿Qué problema resuelvas con tu investigación?

|  |
| --- |
| *Ej: El dolor lumbar es una de las causas más recurrentes de los pacientes en centros de fisioterapia. El tratamiento estándar actual no es efectivo para todo tipo de pacientes. En esta investigación se plantea uno de novedoso para mejorar a todo tipo de pacientes más rápidamente.*  ¡Tú turno!   1. Actualmente por temas relacionados al COVID, tenemos clientes de los préstamos que se han acogido al subsidio y otros que si han seguido pagando. Tenemos un grupo llamado modificado y no modificado y se plantea responder si los clientes que están bajo la categoría modificados tienen relación con áreas o actividades económicas más afectadas por el covid. 2. Por temas del Covid y sobre todo al inicio de la pandemia se hizo una clasificación de empresas y sectores económicos asociados a un nivel de riesgo. (bajo-> sectores económicos poco afectados por covid, medio -> sectores con un impacto , alto-> sectores altamente afectados). Esta clasificación se hizo al inicio de la pandemia y luego de 11 meses es ideal saber o validar si el sector quedo bien clasificado en función de la mora que han mostrado. |
|  |

¿Por qué es importante tu investigación hoy?

|  |
| --- |
| *Ej: el dolor lumbar es la causa más recurrente en centros de fisioterapia. Mejorar el tratamiento de esta ayudará a muchas personas. Demostrar que el nuevo tratamiento funciona mejor aportará claridad al resto de la comunidad.*  ¡Tú turno!   1. Responder, esta investigación nos permitirá determinar si debemos cambiar políticas de crédito a ciertas áreas o industrias y esto puede reducir los costes por provisión de préstamos malos. 2. Nos permitirá saber si la clasificación estuvo bien hecha y de no estarlo podemos reclasificar los sectores de acuerdo a la evidencia de la mora 11 meses después. |
|  |

¿Cómo va a ayudar tu investigación a mejorar la sociedad? O dicho de otro modo ¿cómo mejorará la vida al resto de personas?

|  |
| --- |
| *Ej: en esta investigación se plantea un nuevo tratamiento que se plantea que muchos de los pacientes con dolor lumbar van a mejorar más rápidamente que con un tratamiento convencional. En resumen, un tratamiento más efectivo y rápido para la gran mayoría de pacientes.*  ¡Tú turno!   1. A priori creo que si el banco es capaz de identificar si hay clientes de sectores mas afectados que otros, podemos controlar mejor el riesgo y evitar pérdidas que puedan poner en peligro los puestos de trabajo de nuestros colaboradores. 2. Esta revisión de los sectores y su nivel de riesgo por afectación puede ayudar a flexibilizar políticas de créditos a sectores que por políticas fueron cerrados para obtener créditos. |

¡Bien hecho!

Estás respondiendo preguntas muy importantes.

Esto te dará claridad para definir el análisis de datos.

# EL OBJETIVO

Antes de trabajar los objetivos quiero que expliques muy resumido en qué consiste tu investigación. Escribe un resumen muy rápido de lo que tienes pensado hacer (o has hecho)

|  |
| --- |
| *Ej: en esta investigación se ha tratado dos grupos de pacientes. El primero de ellos formado por 100 pacientes dónde se le ha trabajado un tratamiento convencional, es el grupo control. Y el segundo grupo, el grupo de experimentación. En este último grupo se ha utilizado un tratamiento novedoso. Se han recogido diferentes características de los pacientes para poder comparar los dos tratamientos*  ¡Tú turno!   1. Actualmente tenemos alimentado una base de datos donde tenemos que clientes han sido clasificados como Modificados, cuales No son modificados y a que industria pertenecen. 2. Tenemos una base de datos con los sectores, el nivel de riesgo asignado y contamos con la mora mes a mes que cada sector ha mostrado durante 11 meses.. |

Teniendo en cuenta el contexto y la respuesta anterior lista al menos 3 objetivos (qué quieres conseguir) de tu estudio e identifica los objetivos con (\*) = Poco importante (\*\*) = Importante (\*\*\*) = Muy importante

|  |
| --- |
| *Ej: Los objetivos son:*   * *(\*\*\*) Demostrar que el tratamiento de investigación es más efectivo que el tratamiento estándar.* * *(\*\*) Observar si el nuevo tratamiento mejora igual o diferente a los pacientes según el tipo de paciente* * *(\*) Observar si la pierna corta aparente influye en una lesión de mayor gravedad y si el nuevo tratamiento ayuda más que el convencional*   ¡Tú turno!   1. (\*\*\*) Determinar si existen alguna asociación entre los sectores económicos y los préstamos que han solicitado alivios y han sido modificados. 2. (\*\*\*) Determinar si la clasificación de riesgo asignado a cada sector al inicio de la pandemia están correctamente clasificados de cara al comportamiento actual de la morosidad. 3. (\*\*\*)Ver la proporción que que pueden haber de créditos con un cliente, con uno mas clientes y cuantos clientes pertenecen a sectores de riesgo criticos |

Escoge el objetivo más importante y a la vez más sencillo de lograr. Es muy importante que para empezar en el análisis de datos trabajes con objetivos sencillos. Después ya podrás ir complicando los resultados de tu estudio.

Simplifica el objetivo que has escogido en otras palabras según las características que has medido en tu base de datos.

|  |
| --- |
| *Ej: El objetivo seleccionado es : Demostrar que el tratamiento de investigación es más efectico que el tratamiento estándar*  *Este objetivo es un poco genérico. Podríamos simplificarlo así:*   * *¿El tratamiento de investigación mejora los pacientes más rápidamente que el convencional? Podemos observar el estado de la lesión antes del tratamiento y después de un mes.* * *¿El tratamiento de investigación mejora los pacientes más rápidamente que el convencional en todos los niveles de la lesión? Podemos observar la lesión antes y después de un mes según los diferentes grados de lesión*   ¡Tú turno!   1. Para mi caso de estudio el Objetivo seleccionado es muy sencillo y definido por lo que no tiene sub objetivos. En ambos casos |

# LAS VARIABLES

Define las variables con tus palabras y clasifícalas según el tipo. Esto te ayudará mucho para situar un mejor análisis. Además te forzará a entender todas las variables sin excepción ;)

Los tipos de variables son los siguientes:

* Cuantitativa Discreta *(Ej: Edad)*
* Cuantitativa Continua (Ej: altura en metros)
* Cualitativa Nominal (binaria o politómica) *(Ej: Continente - 1=Asia, 2=África, 3=Europa, 4=América, 5=Oceania)*
* Cualitativa Ordinal *(Ej: escala de dolor (1 = no dolor, 2= dolor leve, 3=dolor soportable , 4 = dolor intenso 5 = dolor insoportable)*

*Aquí el ejemplo:*

|  |  |  |
| --- | --- | --- |
| **Nombre** | **Tipo** | **¿Qué característica observa?** |
| *Id Paciente* | *Cuantitativa Discreta* | *Identificador del Paciente. Número identificado.* |
| *Sexo* | *Cualitativa Nominal* | *Sexo (0 = Hombre 1 = Mujer)* |
| *Edad* | *Cuantitativa Discreta* | *Edad del Paciente* |
| *Peso* | *Cuantitativa Continua* | *Peso del paciente en kg* |
| *Altura* | *Cuantitativa Continua* | *Altura en metros* |
| *LC* | *Cualitativa Nominal* | *Tipo de Lumbociática (0 = NO Lumbociática, 1 = Lumbociática Derecha y 2 = Lumbociática Izquierda)* |
| *PC* | *Cualitativa Nominal* | *Pierna corta aparente ( 0 = No Pierna Corta, 1 = Pierna Corta Derecha, 2 = Pierna Corta Izquierda)* |
| *NHD* | *Cualitativa Ordinal* | *Número hernia discal (0 = Ninguna, 1 = Una, 2 = Dos)* |
| *ODIMes0* | *Cuantitativa Continua* | *Índice de Discapacidad de Oswestry Mes0 (Porcentaje Antes Tratamiento)* |
| *ODIMes1* | *Cuantitativa Continua* | *Índice de Discapacidad de Oswestry Mes1 (Porcentaje Después Tratamiento)* |
| *Grupo* | *Cualitativa Nominal* | *Tipo de Tratamiento (0 = Convencional, 1 = Investigación)* |
| *LF\_Mes0* | *Cualitativa Ordinal* | *Límite Funcional Mes0 (0 = Mínima, 1 = Moderada, 2 = Intensa, 3 = Discapacidad, 4 = Máxima)* |
| *LF\_Mes1* | *Cualitativa Ordinal* | *Límite Funcional Mes1 (0 = Mínima, 1 = Moderada, 2 = Intensa, 3 = Discapacidad, 4 = Máxima)* |
| *IMC\_Grupos* | *Cualitativa Ordinal* | *Estados Nutricional (0 = desnutrido, 1 = Normal, 2 = Sobrepeso, 3 = Obesidad, 4 = Obesidad Mórbida)* |

*Nota: para las variables cuantitativas indica si son de contaje y las unidades de medidas. Para las variables cualitativas indica los grupos*

¡Tú turno!

Rellena la tabla con tus variables:

|  |  |  |
| --- | --- | --- |
| **Nombre** | **Tipo** | **¿Qué característica observa?** |
| *Fecha Corte* | *Serie Temporal* | *Fecha de la observación(Mensual)* |
| *Tipo de Prestamo* | *Cualitativa Categórica Nominal* | *Si el préstamo es Hipotecario o de Consumo* |
| *Nivel de Mora o delinquency* | *Cualitativa Categórica Ordinal* | *En que nivel de mora esta el prestamo* |
| *Empresa del 1er Deudor* | *Cuantitativa – Discreta* | *En que empresa labora el cliente principal del credito* |
| *Sector* | *Cuantitativa – Discreta – Categorica* | *En que sector económico se encuentra la empresa* |
| *Nivel de Riesgo del Sector* | *Cuantitativa – Discreta – Categorica* | *A que nivel de riesgo se le asigno el sector al inicio de la pandemia según se consideró sería impactada (Ato-Moderado-Bajo)* |
| *Categoria* | *Cualitativa Ordinal* | *Indica si el préstamo esta bajo la categoría de Modificado o No Modificado* |
| *Saldo* | *Continua* | *Saldo del Prestamo* |
| *Tipo Mora* | *Categorica - Ordinal* | *Vigente, Moroso y Vencido* |
| *Empresa Segundo Deudor* | *Cuantitiva Discreta* | *En que empresa labora el cliente principal del credito* |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Ahora es el turno de traducir el objetivo que te has planteado con las variables que has medido.

¡**Este paso es muy importante!**

¡Let’s go!

|  |
| --- |
| *Ej: El ODI en el mes0 y el ODI en el mes 1 ha disminuyendo más en el grupo de tratamiento de Investigación que en el grupo control*  *Y también,*  *Replicar esta pregunta por grupos de pacientes en LF\_Mes0. Ver si el tratamiento de investigación es igual o menos efectivo para estos casos.*  ¡Tú turno!   * Ver el promedio saldos vencidos al inicio de la pandemia vs el mes de Dic de 2020 agrupados por Tipo de Préstamos, Sector de la Economía * Ver el promedio de los saldos vencidos agrupados por tipo de riesgo al inicio y al final de la pandemia * Comparar los créditos que tienen mas de un cliente como están agrupados de los sectores de riesgo. |

¡Felicidades! ¡Has avanzado mucho!

Las variables son la materia prima del análisis de datos de éxito. ¡No lo olvides!

# DEFINIENDO EL ANÁLISIS

Ahora que ya has definido el objetivo de tu análisis de datos y sabes el tipo de variables que tienes, es momento de darle forma al análisis.

Con la información de los problemas tipo lo que has visto en el vídeo selecciona el tipo de problema que vas a trabajar.

* Comparación de grupos
* Comparación de proporciones
* Asociación de proporciones
* Análisis de correlación
* Modelos matemáticos
  + Explicar un fenómeno real
  + Predicción
* Técnicas avanzadas (machine learning)
  + Clustering (clasificación)
  + Reducción dimensional (PCA)
  + Modelos avanzados (redes neuronales)
  + Series temporales

¿Cuál de los problemas tipo crees que vas a necesitar para tu objetivo que has seleccionado?

|  |
| --- |
| *Ej: Comparación de grupos*  ¡Identifica el problema tipo!  Comparación de Grupos  Comparación de Proporciones  Asociación de Proporciones  Correlacion entre Saldos Vencidos y Sectores |

¿Qué variables son las que vas a involucrar en este problema tipo?

|  |
| --- |
| *Ej: la variable nominal grupo para identificar el grupo de pacientes y el ODI en els mes 0 y mes 1 para identificar el grado de la lesión antes y después. La variable LF del mes 0 también nos servirá para replicar el análisis para los diferentes grados de lesión (por grupos de lesión)*  ¡Lista las variables que vas a involucrar!  Fecha.  Sector  Categoria  Delicuency  Nivel de Riesgo  Saldo  Con un cliente, con dos clientes |

La variable respuesta es la variable que quieres estudiar. En tu caso, ¿cuál es esta?

|  |
| --- |
| *Ej: VARIABLE RESPUESTA: ODI en el mes 0 y en el mes 1*  ¡Tú turno!  VARIABLE RESPUESTA:  Morosidad al Inicio del Periodo vs Morosidad al Final del Periodo  Cantidad de préstamos modificados por sector y riesgo |

Y las variables explicativas son las características que se involucran en tu problema tipo y que influyen en la variable respuesta. ¿Cuáles son?

|  |
| --- |
| *Ej: VARIABLES EXPLICATIVAS:*   * *variables grupo de tratamiento* * *la variable ordinal LF en el mes 0 para replicar el análisis por grupos de lesión lumbar*   ¡Tú turno!  VARIABLES EXPLICATIVAS  Tipo de Producto o Préstamo  Fecha  Sector Económico  Categoría  Nivel de Riesgo  Saldo |

¡Bien! Estás avanzando mucho.

Ahora tienes el plan de acción casi definido. Solo falta saber qué quieres observar en la etapa de descripción y qué preguntas quieres responder en la etapa de análisis.

¡A por ello!

# LA METODOLOGÍA KILLER

En el vídeo ya has visto los pasos necesarios para analizar los datos de tu tesis. Estos pasos se repiten una y otra vez.

Voy a asumir que has logrado obtener una base de datos. Has limpiado los datos y ahora quieres analizar estos datos para poder lograr resultados impactantes que demuestren que puedes resolver el problema que has planteado al inicio de la hoja de trabajo.

Principalmente vas a pasar por dos etapas:

* La descripción: traducir tu tabla de datos a gráficos que puedas interpretar
* El análisis: definir y responder a las preguntas para poder lograr el objetivo/s que te has marcado en esta hoja de trabajo

## LA DESCRIPCIÓN

Lista qué aspectos quieres ver en tus datos. Focaliza tu atención en las variables. Qué preguntas tienes y quieres resolver con rapidez. La descripción es esencial para contestar a la mayoría de preguntas que te inquietan. Idealmente estas preguntas deben estar relacionadas con el objetivo que has definido.

¡Adelante!

Lista las preguntas que quieres responder. Utiliza las variables dentro de las preguntas. La estadística trabaja con variables ;)

|  |
| --- |
| *Ej: Estas son algunas característica que como investigador de este estudio me gustaría observar:*   * *¿El ODI en el mes 0 es igual para los pacientes del grupo control que en el de investigación?* * *¿El ODI en el mes 1 es diferente para los pacientes del grupo control que en el de investigación?* * *Responder a las respuestas anteriores por grupos de pacientes según al LF en el mes 0* * *¿El ODI0 es mayor con pacientes con hernia discal?* * *¿Tener pierna corta aparente influye en tener un ODI mayor en el mes0?*   ¡Te toca! |

## EL ANÁLISIS INFERENCIAL

Este ejercicio nos permitirá pasar de un objetivo a preguntas de investigación con lenguaje en clave de variables. Ahora lista al menos dos preguntas de investigación. Estás serán el embrión de tus preguntas para los contraste de hipótesis.

Guíate del ejemplo para lograrlo. Utiliza las variables. La estadística trata de responder preguntas teniendo en cuenta variables de la tabla de datos.

¡A por ello!

Seguramente en el ejercicio anterior ya has definido preguntas de investigación muy interesantes. ¡Cópialas aquí!

|  |
| --- |
| *Ej: copio las preguntas más relevantes para lograr el objetivo de la investigación*   * *¿El ODI en el mes 0 es igual para los pacientes del grupo control que en el de investigación?* * *¿El ODI en el mes 1 es diferente para los pacientes del grupo control que en el de investigación?*   ¡Tú turno! |

# ¡ENHORABUENA!

¡Has logrado tu plan de Acción: “Análisis de Datos de Éxito”!

Has conseguido desgranar tu investigación en pequeños objetivos. Seguramente sientas más foco y menos lío.

Estos pequeños objetivos puedes alcanzarlos con la ayuda del análisis de datos.

Te aconsejo que repitas el pequeño ejercicio “Definiendo el Análisis” para los 5 objetivos que te has marcado en tu investigación.

Esto te dará mucha claridad para decidir qué test y técnicas tienes que utilizar para analizar tus datos.

Ya te dije al principio que no hace falta saber de estadística para lograr terminar la hoja de trabajo. ¡Lo has hecho simplemente siguiendo una hoja de ruta!

¡El siguiente paso es pasar a la acción con herramientas estadísticas! Y de esta manera describir (crear gráficos) y analizar (inferir).

Para ello te he preparado un bonus de herramientas dónde puedes encontrar las opciones para crear gráficos potentes y calcular test estadísticos que ¡te van acelerar tu tesis hacia el éxito!

¡Nos vemos en el bonus!

PD: tómate una caña, un zumo de zanahorias o lo que más te apetezca. Has logrado algo importante para tu tesis ;)

¡1 fuerte abrazo!



Ayudo a estudiantes de doctorado e investigadores a analizar los datos de sus tesis para que puedan presentarlas con éxito ante el tribunal y consigan el reconocimiento y el trabajo investigador de sus sueños

Jordi Ollé

Por favor dime qué te ha parecido esta guía contándome tu experiencia a [jordi@conceptosclaros.com](mailto:jordi@conceptosclaros.com) . ¡Espero haberte ayudado!