Preguntas SAS.

NOTA: Las respuestas marcadas en rojo es lo registrado como incorrecto **(Favor de cambiar si se detectan inconsistencias)**

* DM007.Son dos funciones de SAS que se usan para calcular **estadísticas descriptivas**
  + **PROC UNIVARIATE y PROC MEANS**
* DM013. El area bajo la curva ROC equivale a esta estadistica
  + **Estadística Wilcoxon**
* DM007. ¿Que efecto tiene la opción ALPHA en la sentencia MODEL del procedimiento GLM?
  + **Establece que nivel de significancia debe usarse junto con el p-value obtenido en el modelo para rechazar la hipótesis nula**
* DM019. Es posible indicar en la tarea que produce estadísticas descriptivas que la hipótesis nula no sea Mu = 0,sino otro valor:
  + **Verdadero**
* DM019. En una tarea “Linear”, al indicar que la selección de modelo debe de ser por “Backward elimination”, podemos indicar estos niveles de significancia:
  + **De salida solamente**
* DM004. Es el criterio predeterminado de partición “Splitting Rule” en el nodo “Decision Tree”, para las variables objetivo de tipo binario:
  + **Prob Chi-square test**
* DM001. Un árbol puede construirse interactivamente:
  + **...Oprimiendo el botón de “elipsis” (puntos suspensivos) a la derecha de la propiedad “interactive” al seleccionar ese nodo**
* DM013. Un buen analista debe al menos tener conocimiento de …
  + **El dominio del negocio en que trabaja, los ambientes de datos que alli se utilizan y metodos analiticos relevantes**
* DM019. ¿Por qué es importante revisar si existen valores extremos antes de utilizar una variable para una regresión?
  + **Porque los valores extremos suelen provocar influencia desmedida sobre los parámetros del modelo**
* DM001. Una variable cuyos niveles sean “03, Benito Juarez” …. puede configurarse mejor con el tipo:
  + **Nominal**
* DM001. Una variable cuyos niveles sean “Nulo”, “Chico”, “Mediano” y “Grande” puede configurarse mejor con el tipo:
  + **Ordinal**
* DM004. Cuando esta propiedad es “Yes” el nodo “Decision Tree” crea una nueva variable indicando el nodo hoja en el que queda catalogado cada registro
  + Variable Selection
* DM019¿Que correlaciones podemos solicitar en una tarea Correlations?
  + **Pearson, Hoeffding, Kendal, Spearman**
* DM022. ¿Que debemos reportar de una segmentación para que podamos reproducir en otras plataformas, o frente a datos?
  + **Las variables usadas, el tratamiento que se les dio, y la ubicación de los centroides de cada cluster**
* DM004. El nodo “Decision Tree” solo es especialmente apto para trabajar con variables categoricas
  + **Falso**
* DM022, Un modelo obtenido por K-Means y directamente usado para que, por sus números identificadores, se toman acciones para cada cluster, ¿que puede pasar cuando se repite el modelado?
  + **Que los números de cluster nuevos no coincidan con los números de la versión anterior**
* DM019. Pueden reportarse los cuantiles desde esta tarea:
  + **Analyze / Distribution Analysis**
* TS001 El suavizamiento exponencial es una operación accesible desde…
  + **Proc. ESM**
* DM015 Aunque se configure a SAS Entreprise Miner para formar múltiples ramas por nodo, los reportes gráficos siempre muestran ramificaciones por pares (bifurcaciones)

DM016 Aunque se configure a SAS Entreprise Miner para formar múltiples ramas por nodo, los reportes gráficos siempre muestran ramificaciones por pares (bifurcaciones)

* + **Falso. Si se configuran ramificaciones múltiples, así las graficara**
* TS001 Existen mediciones irregulares del consumo de un grupo de clientes. Éstas mediciones en ocasiones incluyen resúmenes, que provienen de gastos transferidos de otras cuentas a la suya. ¿Que procedimiento puede darse a la información previo el análisis?
  + **Todas las anteriores**
* DM010 La correlación de Spearman sirve especialmente si…
  + **La variable predictiva es ordinal y la variable objetivo es binaria**
* DM022, La cantidad de casos en cada segmento puede verse con precisión en la ventana
  + **Mean Statistics**
* DM007 PROC UNIVARIATE solo produce gráficas de texto y se quiere de PROC GPLOT para obtener archivos GIF
  + **Falso, el resultado incluye gráficas GIF si se solicitan**
* DM013 Son técnicas afines a la minería de datos y analytics
  + **Todas las anteriores**
* DM001, Previo a algún nodo de red neuronal, conviene realizar selección de variables
  + **Verdadero**
* DM001 Desarrollar un estudio con una partición de pruebas que no fue obtenida aleatoriamente puede provocar que el modelo…
  + **Reporte un poder predictivo erróneo**
* DM019 Al solicitar un analisi de distribución univariada, podemos solicitar que se reporten los siguientes:
  + **Todas las anteriores**
* DM019 Sas recomienda el uso de gráficas que provee Entreprise Guide para que el analista no necesite de la interpretación de las estadísticas resultantes de los procedimientos
  + **Falso**
  + **Verdadero: la mayoría de las estadísticas importantes pueden verse a simple vista en cualquiera de las gráficas generadas**
  + **Verdadero**
* DM010 Cuando un modelo utiliza una variable de la que se sabe que no tiene poder predictivo )por ejemplo, una variable categórica artificialmente aleatoria), sabemos que ha incurrido en:
  + **Sobreajuste**
* DM022 ¿Como es la codificación de una variable de intervalo, en el proceso que sigue el nodo Clustering?
  + **No es necesario crear dummy variables para una variable de intervalo**
* La letra A del proceso SEMMA corresponde a:
  + **Accionamiento**
* DM022 Dos ejemplos de formulas para calcular la distancia entre observaciones son:
  + **Euclidiana y Manhattan**
* DM016 ¿Como puede obtenerse la asertividad a partir de matriz de confusion?
  + **Dividiendo la suma de la diagional principal entre el total de todas las casillas**
* DM022 La gráfica “Segmente Size” muestra…
  + **Una gráfica de pie con los porcentajes que corresponden a cada segmento**
* TS001 PROC TIMESERIES es capaz de…
  + **Todas las anteriores**
* DM007 Es un ejemplo de variable continua o de intervalo
  + **Los dias transcurridos desde el inciio del contrato**
* DM007 ¿Que indica la sentencia MODEL del procedimiento GLM?
  + **Que se deben considerar para el modelo las variables ahi expresadas con un signo de igual para separar variables efecto de variables predictivas**
* DM013 “La densidad de información es los circuitos integrados de silicon se duplica cada 18 a 24 meses” se conoce como la Ley de…
  + **Moore**
* DM007 PROC UNIVARIATE requiere de PROC SORT simpre que va a ejecutarse
  + **Falso, previo a PROC UNIVARIATE, PROC SORT solo se necesita si se usara una sentencia BY que aplique análisis univariado para cada nivel de esa variable**
* DM010 Al realizar un modelo sin partición de validación ni de pruebas de corre el riesgo de…
  + **Sobreajuste**
* DM010 Para requerir las estadísticas SPEARMAN y Hoefding debemos indicar:
  + **Las opciones SPEARMAN y Hoeffding de las sentencia principal PROC CORR**
* DM014. En un análisis de supervivencia, cuando no ha ocurrido un evento durante el tiempo de estudio, se dice que es un dato
  + **Censurado.**
* DM005 Al producir reglas en inglés para un objetivo binario, ¿cuáles de esos datos se reportan como resultado en cada regla?
  + **Número de nodo, número de observaciones, y porcentajes para cada valor del objetivo**
* DM002 Previo a la creación de clusters mediante K-means, qué tratamiento conviene dar a las variables
  + **Agrupación y binning**
* DM005 El nodo “Decision Tree” con la propiedad “Missing Values” = “Use in search realizará esto:
  + **Utilizará los valores como un nivel más de esa variable**
* DM005 Es el criterio predeterminado de partición (“Splitting Rule”) en el nodo “Decision Tree” para las variables objetivo del intervalo.
  + **Prob Chi-square test**
* DM005 En el resultado de un nodo “Decision Tree” una variable puede aparecer varias veces, en varios niveles, a menos que indiquemos:
  + **La propiedad “Use Input Once” = YES**
* DM002 Un resultado de segmentación ha producido agrupaciones que parecen rombos y cubos. ¿Qué tipo de distancia es el que más probablemente se usó?
  + **Euclidiana**
* DM017 Siempre debemos pedir el reporte de reglas a manera de texto para implementar un árbol de código en SAS
  + **Verdadero: la generación de código solo funciona desde la ventana de English Rules**
* DM008 ¿Qué sentencia indica a PROC GLM que se desea una regresión polinomial multivariada?
  + **model y = x x\*x z;**
* DM017 El histograma de cualquiera de los nodos de un árbol es igual al histograma del nodo raíz a escala
  + **Falso, la distribución en un nodo cualquiera no tiene que ser igual al de la raíz**
* DM008 ¿Cuándo se utiliza la opción OUTPUT de las sentencias ODS?
  + **Cuando se desea generar tablas SAS con los resultados que reporta un procedimiento**
* DM008 ¿Qué efecto tiene la opción CLI en la sentencia MODEL del procedimiento GLM?
  + **Produce y reporta los intervalos de confianza**
* DM005 Un espacio de atributos formado por variables categóricas y ordinales no puede analizarse mediante árboles
  + **Falso. El árbol puede formar regiones usando variables categóricas**
* DM011 Si se desea que PROC REG seleccione las variables para integrarlas paso por paso al modelo, es necesario usar la opción SELECTION=STEPWISE, FORWARD o BACKWARD.
  + **Verdadero**
* DM008 La diferencia principal entre un diseño hecho con PROC GLM y otro hecho con PROC REG está en…
  + **PROC GLM tratará a las variables del modelo cono categorías, PROC REG las tratará como variables continuas.**
* DM002 Al usar Chi2 en el nodo “Variable Selection” para seleccionar variables categóricas, ¿cómo se detectan las interacciones?
  + El nodo realiza un árbol y transmiten con el rol de “Input” al siguiente nodo las variables seleccionadas como predictores por este árbol
* DM020 Dos tareas que trabajan con reportes de percentiles son…
  + **Analyze/Distribution Analysis and Analyse/Summary Statistics**
* DM008 ¿Qué efecto tiene las sentencia CLASS dentro del procedimietno PROC GLM?
  + **Subdividir el procedimiento, generando diferentes modelos para cada clase**
* DM014 La selección de variables, los análisis univariados y los reportes de asociación entre variables son ejemplos de procesos…
  + **De exploración, acorde a SEMMA**
* DM005 La estadística LogWorth (-Lop(p)) usada en el árbol interactivo está principalmente ligada a:
  + **La entropía**
* DM002 El nodo “Neural Network” permite que se configuren qué aspectos generales de la red neuronal:
  + Direct Connection, Number of Hidden Units, Sigmoidal Activation
* DM005 Es posible modificar interactivamente un árbol que ya ha sido entrenado automáticamente
  + **Verdadero**
* DM005 En los nodos “Decision Tree”, establecer la propiedad “Number of Surrogate Rules” mayor a cero ayuda a resolver qué problema:
  + **Variables con valores faltantes**
* DM008 Cuando ejecutamos ODS TRACE ON; ODS LISTING CLOSE; previo a un procedimiento, ¿qué reporta SAS?
  + El resultado normal del PROC que se haya ejecutado, pero son listas de variables.
* TS002 Esta opción indica a PROC ARIMA que debe analizar la serie de la variable V1, obteniendo diferencias con el lapso 1 y lapso 12.
  + **IDENTIFY VAR V1(1,12);**
* DM002 El nodo “Reggression” requiere que la variable objetivo sea del tipo:
  + **Intervalo, binario u ordinal**
* DM002 Cuando el objetivo es una variable categórica con 3 niveles, ¿hasta cuántos estudios de regresión logística potencialmente pueden hacerse?
  + **Cinco**
* DM014 Este nodo se puede configurar para realizar un muestreo aleatorio estratificado, tomando 100% del nivel más raro, y una cantidad igual del otro nivel de la variable objetivo
  + **Sample**
* DM011 Si se desea que PROG REG considere a todas las variables para incluirlas en el modelo, es necesario usar la opción SELECTION = RSQUARE, ADJRSQ, CP, o NONE.
  + Falso
* DM020 ¿En qué condiciones se puede preferir Enterprise Guide sobre Enterprise Miner?
  + **Cuando se desea un estudio y un reporte que ya están preconfigurados en Enterprise Guide**
* TS002 PROC ARIMA puede ejecutarse sin intervención del usuario
  + Verdadero: el algoritmo calcula los mejores parámetros de lapso (lag).
* TS002 Se ha recabado información periódicamente, pero debido a asuetos y manejos del tiempo, la información tiene algunas mediciones faltantes a lo largo de una serie de tiempo. ¿Qué proceso puede ayudar?
  + **PROC EXPAND**
* DM011 Para elegir entre dos modelos de regresión, es mejor tomar el que
  + **cumple con los criterios Mallows Cp y Hocking’s, tiene la mayor R2 ajustada y el menor error cuadrático promedio**
* DM011 Cuando hay valores extremos presentes en la base de datos, ¿cómo puede afectar a la regresión logística?
  + **Los valores extremos provocan leverage (apalancamiento) que desvían el cálculo de la ubicación óptima del modelo**
* DM008 Si ODS reporta un elemento con nombre BasicMeasures y con ruta Univariate.weight.BasicMeasures, ¿cómo podemos invocar justo ese resultado?
  + ODS select data = Univariate.weight.BasicMeasures print;
* DM011 ¿Qué reporta la opción VIF en PROC REG?
  + **El factor de inflación de la varianza por cada variable**
* DM011 El reporte obtenido por PROC REG muestra una estadística F significativa, pero ningún término significativo. ¿Qué problema es probable que haya ocurrido?
  + **Colinealidad**
* DM017 Al preparar las variables para une studio, fácilmente puede perderse información útil para un modelo de árbol
  + Debe cuidarse que las agrupaciones de variables con demasiados niveles no sobre-simplifiquen y pierdan información
* DM002 Un resultado de clustering se consideró equivalente a un estudio de regresión logística, ya que se hallaron tantos clusters como clases que se quería encontrar y coinciden aproximadamente las clases con -
  + Cada re-ejecución de k-means cambia las ubicaciones de los centroides y las distribuciones halladas
  + La importancia de las variables no será reportada
* DM011 El problema de pseudo-replicación puede ocurrir si hacemos un muestreo…
  + **Con repetición.**
* TS001 Son tipos de suavizamiento exponencial que pueden probarse para buscar el mejor ajuste de serie de tiempo
  + Simple, Double, Linear y Winters
* DM016 Un árbol forma sub-segmentos que además de coincidir con la variable objetivo, puede estar relacionados con otras variables ¿Cómo puede visualizarse fácilmente?
  + Con un nodo Stat Explore
  + Con un nodo Profilin
* DM022 La gráfica “Segment Size” muestra..
  + **Una gráfica de pie con los porcentajes que corresponden a cada segmento**
* DM007 Para producir polígonos de frecuencias en PROC UNIVARIATE ¿Qué sentencia (statement) debe añadirse?
  + **La sentencia HISTOGRAM**
* DM019 Si estoy trabajando sobre una tabla que contiene un resumen de la tabla de datos original, donde una variable indica cuántas veces hubo un reglón con justo esos mismos datos (el conteo de ocasiones en que ocurrió esa observación) ¿dónde debe indicarse en la tarea Analyze/ Multivariate/ Correlations?
  + Dando a esa variable el rol “Relative Weight”
* DM004 El árbol de clasificación subdivide al espacio de atributos ¿En que forma?
  + **Todas las anteriores**
* DM001 Los siguientes son resultados automaticos que siempre estan presentes en el resultado del nodo de comparación de modelos, cuando la variable objetivo es binaria.
  + Curva ROC, gráfica Kolmogorov-Smirnov, coeficiente de la partición de pruebas e indicador del punto de corte óptimo para el mejor modelo
  + Curva ROC, respuesta capturada acumulativa, gráfica de estadística Kolmogorov-Smirnov
* DM004 ¿Cual es la diferencia al cambiar la propiedad Assessment Measure de “Average Square Error” a “Decision”?
  + **Que el árbol se podará según ofrezca mayores ganancias/ menores pérdidas acorde a la matriz de ganancias que debe haberse configurado**
* DM019 La regresión logística requiere que las variables predictivas sean de tipo numérico, de intervalo.
  + **Falso: pueden generarse varias variables “dummy” por el mismo algoritmo**
* DM013 El nodo de tipo input puede usarse para indicar la matriz de decisiones
  + **Verdadero**
* DM013 Para establecer las propiedades de los nodos con que integremos el estudio, debemos…
  + **Seleccionar el nodo y ver el panel de propiedades**
* DM010 PROC UNIVARIATE permite la creación de un dataset que contenga las medianas de algunas variables.
  + **Verdadero**
* DM016 En la matriz de confusión, podemos encontrar los falsos positivos en donde..
  + **El valor predicho es 1 y el valor real es 0**
* DM019 ¿Que gráfica, que estudio y que proceso pueden indicar que dos grupos difieren en el importe mensual de consumo?
  + **Una gráfica de box y un estudio de anova provenientes de PROC GLM**
* DM013 los roles que pueden tomar una variable en SAS Enterprise Miner son…
  + **ID, Time ID, Input, Target, Rejected, Text, Frequency, Hidden, Cost**
* DM013 Para establecer la “semilla” de los números aleatorios y poder hacer repetibles distintas ejecuciones de un diagrama, podemos hacerlo desde la ventana:
  + **Preferences**
* TS001 El proceso AUTOREG añade pasos de análisis para identificar y medir autocorrelación considerando heterocedasticidad ¿Qué métodos puede utilizar?
  + **Todas las anteriores**
* DM001 El nodo “Regression” realizará automáticamente los siguientes pasos:
  + Seleccionar las variables de mayor R2 como entradas(...) “Forward”, “Backward” o “Stepwise” segun se le haya configurado, reportando el mejor modelo
* DM013 En caso en que hayamos configurado una matriz de ganancias apropiada ¿donde podemos ver gráficas con la ganancia esperada para los distintos percentiles del modelo seleccionado?
  + **En el nodo Assessment, de modo predeterminado**

**BUSINESS ANALYST - ZIP**

* DA002 Si no se cuenta con información para un análisis, ¿es posible desarrollar un modelo?
  + **Sí, se pueden hacer modelos como tal, por medio de datos simulados, pero su uso y aportación será muy distinta a la del modelado predictivo típico.**
* TA002 Un consultor ha tomado una serie de tiempo complicada (con periodicidad trimestral en vez de mensual, y datos faltantes en algunos periodos), y ha mejorado al modelo previo en todas las estadísticas reportadas p-
  + **Si ha calculado valores faltantes y alterado la periodicidad mediante un modelo, de modo que las estadísticas de calidad sólo confirmen alta congruencia con ese mismo modelo.**
* DA011 Este es un ejemplo de regresión inapropiada que puede descubrirse al graficar los datos y los errores
  + **Cuando se forma una relación curvilínea entre las variables y el objetivo**
* DA008 PROC GLM utiliza la sentencia MODEL a = b c; para indicar…
  + **Que el modelo completo a considerar es A = BaseLevel+CoefA + CoefB + Error**
* DA008 ¿Qué debe hacerse previo al inicio de un estudio?
  + **Todas las anteriores**
  + **Documentar las preguntas de interés**
  + **Definir la población bajo estudio**
  + **Determinar el protocolo de recolección de datos**
* DA008 Los intervalos de 95% de confianza para los coeficientes (Betas) de un modelo de regresión fueron entregados tarde. ¿Qué podría ser más preocupante al ver estos intervalos?
  + Ver que alguno de los intervalos incluye al cero
* DA008 ¿Qué procedimientos son capaces de calcular el número de casos, promedio, suma, mínimo, máximo, percentiles y rango intercuartil?
  + **Solo PROC UNIVARIATE**
* DA011 Una gráfica Logit Plot con la variable predictiva en el eje horizontal y logit de la probabilidad de un evento en el eje vertical forma una distribución lineal diagonal ascendente. ¿Qué podemos concluir de la variable pr-
  + **Qué no aporta por encima del azar**
* DA002 Un resultado de segmentación ha producido aportaciones que parecen globos, ¿qué tipo de distancia es el que más probablemente se usó?
  + **Lyapunov**
* DA011 Un software para manejo de versiones de anuncios por internet es capaz de realizar miles de experimentos por semana. Se han obtenido muchos modelos bien manejados pero las campañas en donde se han aplicado-
  + **Al realizar miles de estudios del tipo ANOCA; con la usual alpha de 0.05, se tendrá a la larga un cinco por ciento de errores en las recomendaciones. La gran población de estudios facilita además el problema de la selección de resultados favorables.**
* DA020 Se indica en estos casos no usar la estadística Chi-cuadrada
  + **Cuando al menos 20% de las celdas de la tabla formada por las variables tiene menos de 5 elementos**
* DA008 Una población teórica se distingue porque
  + **Constantemente cambia**
* DA014 Si una base de datos ha sido analizada para satisfacer un requerimiento analítico, pero no se logra llegar al poder predictivo que justifica al caso de negocios, ¿qué puede proponerse a los stakeholders tal que fuera-
  + **La búsqueda de mejores algoritmos**
* DA023 Cuando se tiene una segmentación en la que dos segmentos tienen un perfil prácticamente igual, deben agruparse
  + **No, puesto que provienen de diferentes lugares en el espacio de atributos, se deben considerar distingos y se pueden buscar más variables que los perfilen más claramente**
* DA014 Una red neuronal puede entrenarse mediante aprendizaje supervisado para detectar procesos de fraude, sin la necesidad de que el negocio o el cliente los califique como fraudes
  + **FALSO**
* DA006 Cuando modificamos alfa al entregar un estudio, en efecto, ¿qué estamos haciendo?
  + **Todas las anteriores**
  + **Ajustando el objetivo de negocio para que se adapte a los daños**
  + **Arriesgando más a que el negocio no halle la respuesta esperada**
  + **Arriesgando más a que el modelo tenga menor asertividad**
* DA017 Un árbol nunca requiere de preparación de variables
  + **Falso: el algoritmo requiere que las variables categóricas se conviertan en dummy variables**
  + **Falso: por ejemplo, cuando una variable nominal tiene muchos niveles que pueden agruparse por conocimiento del dominio**
* DA023 El requerimiento de un usuario menciona el trato diferenciado a los segmentos de clientes ¿qué análisis pueden satisfacer este requerimiento?
  + **Segmentación de la base de clientes por métodos de clustering, ANOVA, o árboles de clasificación**
* DA020 El teorema de límite central será más efectivo en un experimento de este tipo:
  + **Un promedio obtenido a 1000 promedios que provengan de muestras de tamaño 5**
* DA008 En una distribución normal, ¿qué porcentaje de la población se encuentra a una distancia máxima de la media de una desviación estándar?
  + **0.68**
* DA002 Un resultado de clustering se consideró equivalente a un estudio de regresión logística, ya que se hallaron tantos clusters como clases se querían encontrar y coinciden aproximadamente:
  + **Todas las anteriores**
  + **Las fronteras de decisión pueden estar en ubicacion sub-optima**
  + **La importancia de las variables no ser reportada**
  + **Las densidades de población no necesariamente tienen que coincidir con las concentraciones de cada clase**
* DA002 Es posible hacer estudios sobre base de datos que no provengan de un data warehouse
  + **Si, SAS permite la lectura de fuentes externas**
* DA005 Solo los resultados de un modelo de árbol pueden exportarse a manera de reglas verbales:
  + **Falso, tambien las regresiones**
* DA002 Los árboles son especialmente aptos para:
  + **Detectar interacciones, ajustarse a no-linealidades y selección de variables**
* DA005 Se ha tomado la decisión de implementar el resultado de un árbol a manera de programa (variable-objetivo binaria). Al producir las reglas, se nota que SAS solo reporta un decimal es
  + **Configurado al nodo “Decision Tree” para que produzca la variable “NODE” seguido de un nodo de “Scoring” que aplique las reglas obtenidas sobre la partición de validación y programación obtenga porcentajes para cada valor de la variable objetivo, con los decimales deseados**
* DA002 Los análisis científicos experimentados difieren de los análisis comerciales oportunistas en:
  + **Todas las anteriores**
  + **El volumen de información usualmente en pequeños análisis científico y enorme en análisi comercial**
  + **La información usualmente esta limpia en análisis experimental, mientras que esta sucia en los análisis comerciales**
  + **La información suele obtenerse de manera controlada en el analisi experimental, mientras que es observada pasivamente en el análisis comercial**
* DA005 Se ha pedido un modelo para formar el menú automático del servicio telefónico a cliente. Se quire clasificar a cada llamada entrante, pero inicialmente se deben
  + **Si, usando un árbol interactivo que comienza separando a las llamadas de emergencia y continúe automáticamente sobre el resto de las llamadas**
* DA017 En el resultado del árbol, los vínculos entre nodos pueden interpretarse como reglas de asociación entre casos
  + **Falso, el diagrama de vínculos es dado por el nodo de associations, no por el de decision tree**
* DA020, Al hacer una tarea Correlations que involucra múltiples variables, solo un par será seleccionado para graficarlo
  + **Falso, siempre se generan todos los apres de variables**
* DA002, Que consecuencias pueden traer el uso multiplicado de casos de una misma clase(practica que aveces se usa cuando existen muy pocos casos de entrenamiento
  + **Sobre-Ajuste**
  + **Exceso en la estimación de ganancias/pérdidas**
* DA014, Los usuarios de finanzas, operaciones y marketing manejan requerimientos similares para contradicciones al final ¿Existen una manera de conciliar sus requerimientos?
  + **Si, se require de juntas que los lleven a unánimemente formular un solo requerimiento, para que todos tengan acuerdos para el mejor funcionamiento de la empresa.**
* DA005 Se ha producido un árbol con buen resultado, pero su tamaño es excesivo (no se ha usado una matriz de ganancias
  + **Asegurarse de que la propiedad subtreeMethod sea “Assessment”; Establecer Maximum Branch y Maximun Depth a los límites requeridos; agrupar manualmente las variables categoricamente árol que tengan demasiados niveles**
* DA002 Que consideraciones se debe hacer sobre las fuentes de datos que no provengan de la propia plataforma SAS?
  + **Se debe tener cuidado sobre los tipos de variables, las llaves que se usaran para uniones y la compatibilidad con los datos que si provengan de la paltaforma**
* DA014 Un área del negocio desea la formación de un modelo predictivo que le permita anticipar si un evento ocurrirá. No se tienen registros del evento, pero sí se cuenta con información de clientes, sus consumos, sus ubicaciones-
  + **Assessment para identificar en qué percentiles hay mayor propensión a hallar el evento buscado**
* DA002 Desarrollar un estudio sin haber separado de antemano a los datos para pruebas, y luego usando parte de los datos de aprendizaje para pruebas, puede conducir a un problema de
  + **Subestimación del poder predictivo**
* DA007 Es un ejemplo de variable discreta.
  + **El número correspondiente a la delegación o municipio, según datos del INEGI (censo)**
* DA007 Si se tiene un dado justo y tras muchos lanzamientos no ha salido 6, ¿cómo refleja SAS el hecho de que en los siguientes tiros aumenta la probabilidad de un 6?
  + **SAS no tiene nada qué ajustar en ese tipo de situación**
* DA022 Un consultor requiere que se le convierta la información a valores numéricos, ya que él afirma que el algoritmo de clustering sólo funciona con información numérica. ¿Qué puede hacerse?
  + **Explicarle que el programa usado convierte las variables categóricas en “dummy variables” interna y automáticamente**
* DA013 ¿Qué gráfica indicaría que los clientes con una característica son más propensos a adquirir un producto?
  + **Dos barras (una para clientes con la característica y otra para los que no) que midan la proporción de los clientes que tienen el producto**
* DA019 Los reportes generados por Enterprise Guide son siempre iguales, por lo que conviene hacer pocos reportes por proyecto
  + **Falso, pueden imprimirse títulos particulares a cada reporte**
* DA016 Se desea diseñar un menú para presentar al cliente que llame por teléfono para solicitar servicios. ¿Cómo puede ayudarle un árbol?
  + **Un árbol interactivo sobre una base que contenga variables para las preguntas potenciales y el servicio que finalmente satisface al cliente como objetivo puede simular la manera en que se desarrollaría el uso del menú. Se puede comparar con árboles obtenidos automáticamente.**
* DA007 ¿Cuál de estos procedimientos puede generar histogramas?
  + **PROC UNIVARIATE**
* TA001 ¿Qué puede esperarse en un correlograma de una serie cíclica?
  + **Que muestre oscilaciones, semejante a una senoidal**
* DA016 Un árbol de decisión puede servir para seleccionar las variables relevantes en un modelo predictivo
  + **Verdadero**
* DA010 Al graficar los residuos de una regresión contra los valores estimados, idealmente esperamos ver…
  + **Puntos aleatorios uniformemente variantes alrededor del eje horizontal**
* DA022 Un modelo de segmentación fue obtenido por una consultora externa y se informó el número de cluster para cada tarjetahabiente y la decisión de negocios a tomar por cada segmento. ¿Qué falta en esa entrega para que el modelo pueda usarse con clientes futuros?
  + **Las variables usadas, el tratamiento que se les dio, y la ubicación de los centroides de cada cluster**
  + **Las variables usadas, el número de elementos que se obtuvieron y los parámetros de la ejecución de k-means**
* DA019 ¿Puede realizarse un estudio sin eliminar los valores extremos?
  + **Si: cuando es justo lo que queremos distinguir y/o cuando se han tomado medidas para asegurar que no hayan observaciones de influencia desmedida**
* DA022 Recursos humanos desea distinguir las contrataciones de alto desempeño y comienza por solicitar una descripción de la población en términos de clusters. ¿Qué variables son más apropiadas para este estudio?
  + **Datos generales de la contratación y métricas de desempeño**
* DA010 Un valor negativo en la correlación Spearman nos indica…
  + **Que conforme una variable asciende la otra desciende**
* DA022 Un propósito común en las actividades de Marketing es dar un tratamiento personalizado al cliente. Un proceso de clustering con perfilado se propone para ese tipo de objetivos, ya que…
  + **El proceso generará grupos más semejantes entre sí**
  + **El proceso es capaz de indicar qué clientes se encuentran en cada grupo**
  + **El proceso puede usarse para explicar qué tienen en común los elementos de un grupo, y en qué difieren de los de otros grupos**
  + **Todas las anteriores**
* DA001 ¿Qué es más deseable y creíble en una curva ROC?
  + **Una curva a manera de arco pronunciado por encima de la diagonal, con un valor ROC mayor a 0.7**
* DA013 En los requerimientos para un desarrollo interno para gestión de campañas de marketing se ha sugerido que el modelo estime el retorno a la inversión. ¿Qué puede hacerse con los resultados directos de un modelo de regresión logística?
  + **Interpretar el resultado del modelo como el logit de la probabilidad esperada, y de ahí calcular la utilidad esperada; usar esto para ordenar la lista de contactos descendentemente y proponer puntos de corte, considerando que cada uno requiere de más inversión**
* DA019 Para configurar un proceso de componentes principales, es necesario…
  + **Declarar cuál es la variable objetivo**
  + **Separar el estudio por segmentos**
  + **Transformar las variables de intervalo a categóricas**
  + **Ninguna de las anteriores**
* DA013 Cuando se aceptan muchas variables para la construcción de un modelo, se agrava el problema de “la maldición de la dimensionalidad”
  + **Verdadero**
* DA019 La gráfica de interacciones entre dos variables explicativas muestra que una de ellas cambia de pendiente. ¿Qué puede estar sucediendo?
  + **Hay información común aportada por ambas variables**
  + **Hay un problema de colinealidad entre ambas variables**
  + **Hay valores faltantes que cambian alguna de las pendientes para alguno de los niveles**
* DA007 Una población es el conjunto de todas las mediciones de interés
  + **Verdadero**
* DA019 Al graficar un histograma, puede configurarse el que se dibuje también una distribución como línea de referencia. ¿De qué distribuciones se dispone?
  + **Normal, Exponencial, Beta**
  + **Log-normal, Gamma, Weibull**
  + **Ambas opciones**
* DA019 La regresión logística puede obtenerse mediante una regresión lineal y posteriormente la adición de una función logit.
  + **Falso: hay una tarea especial para regresión logística**
* DA004 Cuando se desea segmentar a la población ¿qué factor posibilita el uso de un árbol en vez de algoritmos de clustering?
  + **El que exista una variable que califique a cada caso**
* DA019 Las variables categóricas no pueden ser usadas en las tareas Linear, a manera de variable explicativa
  + **Falso: pueden añadirse directamente con el rol “Explanatory Variables”**
* DA022 ¿Qué ventaja puede tener el que se reporte una métrica por segmentos y no para toda la población?
  + **Que no pasarán desapercibidos algunos cambios que puedan haberse dado mayormente en un grupo pequeño**
* DA016 Si el algoritmo de un árbol de decisión usualmente tiene menor desempeño, ¿por qué incluirlo en los análisis?
  + **Puede hallar interacciones con mucha claridad y sencillez**
  + **Puede producir los algoritmos más simples de implementar**
  + **Ambas opciones**
* TA001 Son ejemplos de series de tiempo en forma apropiada para análisis en SAS
  + **Tasa de desempleo en Jalisco, observada mensualmente**
  + **Cuartos ocupados en hoteles de Vancouver, observados diariamente**
  + **Ambas opciones**
* DA019 Si se mantiene colinealidad en las variables de un análisis, ¿qué desventaja tienen los usuarios?
  + **Pueden dar crédito a una variable equivocada, confundiendo su aportación con la de otra muy parecida**
* DA001 ¿Cómo podemos conocer reglas de texto que describan las ramificaciones de un árbol?
  + **Con el comando “Text Rules” en las ventanas de resultados que muestra Enterprise Miner**
* DA004 Se requiere que un diagrama de árbol obtenido en SAS sea graficado en otra plataforma analítica. ¿Qué puede hacerse?
  + **Exportar mediante código a la tabla que define al árbol, para el uso de la otra plataforma**
* DA019 Un usuario ha solicitado tablas de contingencia. Éstas pueden ser generadas desde…
  + **Describe/Contingency Tables**
* DA007 ¿En qué momento conviene definir el propósito del estudio?
  + **Antes de inciar el análisis**
* DA007 La estadística descriptiva se usa para…
  + **Organizar, resumir y concentrarse en las características de los datos**
* DA022 Finanzas ha propuesto un proceso de clustering para generar alertas adecuadas en caso en que una acción tenga comportamiento aberrante. ¿Qué variables son más adecuadas para generar esta segmentación?
  + **El valor de la acción en cada día de la historia disponible, sus tasas de cambio y descriptores de la empresa que la genera**
* DA022 Dos procesos de clustering parecen dar los mismos resultados en los segmentos mayoritarios, pero no en los segmentos más pequeños. ¿Qué puede hacerse?
  + **Descartar los segmentos pequeños como inestables y hacer otro tipo de análisis sobre ellos (quizá outliers)**
* DA019 Mantener variables redundantes puede degradar el análisis, ya que …
  + **Se desestabiliza el cálculo de estadísticas de ajuste y de parámetros**
  + **Se incremente el riesgo de sobreajuste**
  + **Se incrementa el tiempo y el esfuerzo para aplicar el modelo**
  + **Todas las anteriores**
* DA022 Los segmentos obtenidos manualmente han quedado disputados por el resultado de un nodo Clustering. ¿Cómo puede reportarse lo que cada segmentación sugiere?
  + **Mediante los resultados de ejecutar el nodo Segment Profile sobre la segmentación manual y la segmentación del nodo Clustering**
* DA022 Un nodo Clustering ha sido ejecutado y su resultado ha sido aceptado. ¿Cómo puede establecerse en reglas, para no depender de una misma herramienta?
  + **Documentando los centroides**
  + **Documentando la fórmula de distancia utilizada**
  + **Documentando el perfil con que se distinguió a cada segmento formado**
  + **Todas las anteriores**
* DA004 En su uso, ¿qué ventajas ofrecen los modelos de árbol de decisión por encima de otros modelos?
  + **Simplifican los modelos a través de reglas para que cualquiera pueda interpretarlos**
* TA001 Una métrica se encuentra por debajo del valor pronosticado. ¿Qué situaciones podrían señalar que se debe a deficiencias en el pronóstico, y no necesariamente un cambio en la serie?
  + **La falta de estudio de puntos de cambio en la serie**
  + **La serie había sido obtenido muchos períodos antes, y se ha mantenido sin actualizar**
  + **La serie fue aceptada sin comparativo entre modelos, sólo por ser el modelo más claro o popular**
  + **Todas las anteriores**
* DA013 ¿Qué se espera en las estadísticas de la partición de prueba, con respecto a las estadísticas de la partición de entrenamiento?
  + **Que mejoren, señalen mejor ajuste del modelo**
  + **Que se mantengan igual, si es que se ha usado una base de validación**
* DA007 Un modelo de anova puede obtenerse usando PROC GLM, pero ¿pueden indicarse fuentes de variación?
  + **Sí, el sistema siempre usa todas las variables que se indiquen mediante MODEL**
* DA010 Si obtenemos un histograma de los residuos de una regresión, idealmente esperamos ver…
  + **Una distribución normal**
* DA013 En una base financiera, se observa que una proporción de los clientes no cumplen con los criterios que se debían tener para la aceptación de sus créditos. ¿Qué podría provocar esta situación?
  + **Que los importes se encuentren ajustados a valor presente, modificando los montos que se tuvieron en un inicio**
  + **Que la base histórica puede contener cuentas creadas antes de la aplicación de esos criterios**
  + **El que la información se haya procesado de forma incorrecta**
  + **Todas las anteriores**
* DA001 Para elegir el mejor modelo, usando estadísticas que provee el nodo “Model Comparison”, se prefieren modelos con:
  + **Menor “misclassification rate” y menor error cuadrático promedio**
* DA007 Aumentar el tamaño de una muestra siempre asegura mayor precisión en el cálculo de un promedio y su varianza.
  + **Verdadero**
* DA001 El resultado del algoritmo de regresión mejora cuando las variables de intervalo se encuentran normalizadas
  + **Verdadero, sólo si se trata de una regresión lineal**
* DA019 ¿Qué gráfica y qué estudio pueden indicar que dos grupos difieren en el importe mensual de consumo?
  + **Una gráfica de box y un estudio de anova**
* DA022 Dos procesos de clustering parecen dar los mismos resultados, pero el primero ha dado un nombre a los segmentos que no coincide con los del segundo proceso. ¿Qué puede hacerse?
  + **Un proceso que calcule las distancias entre los centroides y re-asigne las etiquetas del segundo proceso al centroide más próximo del primer proceso que no haya sido ya asignado**
* TA001 Puesto que las transacciones no obedecen a un momento exacto, es conveniente trabajarlas mediante resúmenes por unidad de tiempo. ¿Cuál de estas unidades de tiempo representaría la elaboración de un modelo?
  + **Suma de importes en uso de tarjeta de crédito, por semana**
* DA016 Un árbol repite demasiado el uso de una misma variable. ¿Puede mejorarse?
  + **Sí, requiriendo que cada variable se use una sola vez**
  + **Sí, ampliando el número de ramificaciones que sale de cada nodo**
  + **Sí, modificando el árbol a través de uso interactivo**
  + **Todas las anteriores**
* DA007 Un usuario ha aceptado que el mejor modelo es una regresión logística polinomial multivariada, pero ha solicitado que sólo se presenten explicaciones del modelo a través de reportes univariados. ¿Qué puede pasar desapercibido?
  + **Las interacciones y las no-linealidades**
* DA007 Las estadísticas descriptivas usualmente incluyen éstos cálculos.
  + **Número de observaciones, promedio, desviación estándar, máximo y mínimo**
* TA001 El pronóstico de una serie de tiempo calculada hace 10 períodos se ha mantenido(sin actualización) dentro del intervalo de confianza del 95%. ¿Significa esto que no necesita actualización?
  + **No: significa que de actualizarse con datos nuevos, se modificaría poco al pronóstico restante**
* DA001 ¿Qué otro método de minería de datos es famoso y popular, además de SEMMA?
  + **CRISP-DM**
* TA002 ¿Por qué se obtiene un valor muy alto en la gráfica de autocorrelaciones, siempre en el primer valor?
  + **Porque corresponde a un LAG igual a cero**
  + **Porque es la correlación de una variable consigo misma**
  + **Porque alcanza un valor de uno, y uno es el máximo posible que podría ser observado**
  + **Todas las anteriores**
* DA002 ¿Cómo pueden complementarse otros métodos de análisis junto con SEMMA?
  + **SEMMA puede usarse específicamente para la parte de análisis de datos, CRISP-DM para el desarrollo del proyecto completo de minería, y DMAIC para la identificación , definición, implementación, y monitoreo del proyecto**
* DA008 Para ayudar a que un muestreo sea representativo, debe hacerse…
  + **De forma aleatoria y con suficientes observaciones para mostrar la varianza de sus variables**
* DA002 Los análisis científicos experimentales difieren de los análisis comerciales oportunistas en:
  + **El volumen de información usualmente es pequeño para análisis científico y enorme en análisis comercial**
  + **La información usualmente está limpia en análisis experimental, mientras que está sucia en los análisis comerciales**
  + **La información suele obtenerse de manera controlada en el análisis experimental, mientras que es observada pasivamente en el análisis comercial.**
  + **Todas las anteriores**
* DA014 Una base de datos con alta dimensionalidad necesariamente debe tener muchas variables en la tabla SAS que la contiene.
  + **Verdadero: cada grado de libertad representa una dimensión y por lo tanto, una variable**
* DA008 Un modelo no ha pasado la prueba para considerarse significativo. Sin embargo, se desea utilizar al modelo e informar a sus usuarios de qué grado de confiabilidad tiene. Puesto que dichos usuarios están familiarizados con la significancia, ¿qué podemos decirles que justificaría usar el modelo tal como está?
  + **Modificar el alfa a un valor más bajo**
* DA008 Al crear tablas dinámicas con SAS Enterprise Guide, para cada métrica se requiere del uso de una tabla dinámica particular
  + **Falso, Table Analysis permite reportar varias estadísticas por celda**
* DA005 Al diseñar un cuestionario para aplicar a clientes se desea mínima redundancia y un máximo de 5 preguntas, ¿qué propiedades del nodo “Decision Tree” conviene configurar?
  + **Use Input Once = YES, Maximum Branch=5;Maximum Depth =5**
* DA017 En el resultado del árbol, los vínculos entre nodos pueden interpretarse como reglas de asociación entre casos
  + **Falso: el diagrama de vínculos es dado por el nodo de Associations, no por el de Decision Tree**
* DA002 Una variable de fecha es reemplazada por un entero contador de días a partir del 01-Enero-1960. ¿Cómo puede afectar esto?
  + **No afecta, ya que son variables que siempre se excluyen del estudio**
* DM001 Son algunos nodos correspondientes a la etapa “Modify”.
  + **Drop, Impute, Principal Components**
* DM013 La letra “A” del processo SEMMA corresponde al…
  + Análisis
* DM001 En vez de una variable categórica, se genera un indicador numérico (0 ó 1) para usar como variable objetivo numérico en una regresión. ¿Cómo puede afectar esto?
  + Usará por default una regresión logística en vez de lineal
  + Reducirá la memoria que SAS requiere para almacenar la variable
* DM013 La letra “S” del proceso SEMMA corresponde a “Sample”, y por lo tanto, siempre hay que hacer un muestreo aleatorio simple.
  + Verdadero
* DM010 Las variables que serán usadas por PROC CLUSTER para calcular la distancia entre observaciones se indican en la sentencia…
  + **VAR**
* TS001 Para obtener el periodograma que muestra las intensidades de las frecuencias de Fourier observadas en una serie, acudimos a estos procedimientos…
  + PROC SPECTRA
* DM022 El proceso de clustering realiza iteraciones.
  + **Verdadero.**
* DM022 La cantidad de casos en cada segmento puede verse con precisión en la ventana.
  + **Mean Statistics**
* DM004 Cuando esta propiedad es “Yes” el nodo “Decision Tree” asigna el rol “Rejected” a las variables de entrada que no hayan sido usadas por el árbol.
  + Variable selection
* DM004 Cuando esta propiedad es “Yes” el nodo “Decision Tree” crea una nueva variable indicando el nodo hoja en el que queda catalogado cada registro
  + **Leaf Variable**
* DM013 Una variable que en la base de datos es de tipo texto, character o string, en Enterprise Miner puede usarse con cualquiera de los tipos…
  + Intervalo, Nominal, Ordinal.
* DM013 El nodo “Model Comparison” debe integrarse inmediatamente después del nodo “Input” si es que se desean reportes de calidad en el resto de los nodos que integremos en el diagrama.
  + **Falso**
* DM001 El nodo “Regression” realizará automáticamente los siguientes pasos:
  + **Usará todas las variables con rol “Input”, realizará sobre esa base los pasos “Forward”, “Backward” o “Stepwise” según se le haya configurado y reportará el mejor modelo alcanzado.**
* DM007 Son dos funciones de SAS que se usan para calcular estadísticas descriptivas.
  + **PROC UNIVARIATE y PROC. MEANS**
* DM001 El nodo “Regression” requiere de imputación cuando.
  + **Siempre que hayan faltantes**
* DM001 El nodo de Comparación de Modelos obtiene las estadísticas de calidad para las particiones de aprendizaje, validación y pruebas…
  + **Sólo cuando al nodo “Model Comparison” se le ha configurado para realizar las tres particiones**
* DM013 ¿Podemos leer una tabla grande y de inmediato pedir a SAS Enterprise Miner que deseche algunas variables para ahorrar memoria en el procesamiento?
  + **Si, puede usarse la opción DROP del nodo Input.**
* DM016 En los resultados del nodo “Decision Tree” podemos ver las reglas usadas, a manera de texto, solicitando...
  + **Del menú “View”, la opción “Model” y luego “English Rules”**
* DM010 Al ver una gráfica Logit Plot de la variable predictiva vs Logit de la probabilidad de un evento (pensando en una variable auténticamente predictiva con comportamiento lineal). ¿Qué espero encontrarme?
  + **Una distribución de error alrededor de una línea horizontal**
* DM004 La tabla que define un árbol puede ser leída para graficarse en SAS. ¿Por qué proceso?
  + **PROC TREE**
  + **PROC GRAPH\_TREE**
* DM013 La letra “S” del proceso SEMMA corresponde a “Sample”, y por lo tanto, siempre hay que hacer un muestreo aleatorio simple
  + **Verdadero solo cuando se ejecuta a través de SAS Enterprise Miner**
* TS001 Para realizar pronósticos simples de una gran cantidad de variables, se prefiere el procedimiento…
  + **PROC FORECAST**
* DM013 Si obtenemos un histograma de los residuos de una regresión, idealmente esperamos ver…
  + **Una distribución normal**

* DM004 Un diagrama extremadamente corto, pero con todos los controles y reportes de calidad que conviene tener, estaría formado por: (imaginar datos limpios y único interés en árboles)
  + **Input Data -> Data Partition -> Decision Tree -> Model Comparison.**
* DM019 Para reducir el tiempo de modelado en un árbol, se pueden agrupar manualmente las variables con más niveles.
  + **Verdadero: reducir la cantidad de niveles implica reducir la cantidad de posibles “splits” evaluados.**
* DM001 El siguiente no es un parámetro que puede configurarse en el nodo “Decision Tree”:
  + **Cantidad máxima de casos para considerar una ramificación**
* DM010 Es común usar un Uno o un Cero para indicar cuando un evento ha ocurrido o cuando no ha ocurrido, respectivamente. Esto es para...
  + **Es sólo costumbre, pero actualmente no es indispensable.**
* DM004 ¿Cómo puede aplicarse un algoritmo arbitrario, que preguntando variable a variable, llegue a una calificación, sin que necesariamente haya sido obtenido por un entrenamiento?
  + **No hace sentido aplicar un scoring de esa manera.**