**PROCEDIMIENTO DE ANALISIS**

**HISTORIA**

* El Responsable del área contable necesita saber en que gastaran más por cada trimestre para elaborar nuestro plan de gastos anual

Criterio de aceptación : Nuestro motor de análisis de datos procesa la data recolectada de Contabilidad. Usamos modelos para predecir en que items se gasta más de acuerdo a datos históricos. El nuevo Dashboard mostrará una gráfica con los gastos históricos y la predicción de gastos trimestralmente

**TAREA**

HU5-T2. Análisis exploratorio (para entender tendencias y patrones) y Selección de características y construcción de modelos (entrenar modelos para predecir) Modulo Contabilidad

**DESARROLLO**

Para generar métricas estadísticas sobre los gastos por trimestre, utilizando la estructura de grafos que hemos definido:

* Consulta de Datos Históricos:

Utiliza consultas en Cypher para recuperar los datos históricos de gastos por trimestre. Puedes consultar las facturas y gastos asociados a cada trimestre y sumar los montos correspondientes.

* Análisis de Tendencias:

Utiliza técnicas de análisis de datos para identificar tendencias y patrones en los gastos trimestrales a lo largo del tiempo. Puedes calcular medidas estadísticas como la media, la mediana, la desviación estándar, etc., para entender la distribución de los gastos y detectar cualquier cambio significativo en el tiempo.

* Modelado Predictivo:

Utilizamos modelos predictivos basados en datos históricos para predecir los gastos futuros por trimestre. Puedes utilizar algoritmos de aprendizaje automático como regresión lineal, series temporales, o incluso modelos más avanzados si es necesario.

* Generación de Dashboard:

Utilizamos herramientas de visualización que muestre los gastos históricos y las predicciones de gastos trimestrales.

El dashboard incluye gráficos que muestren los gastos históricos por trimestre y las predicciones futuras basadas en los modelos predictivos.

* Revisión y Ajuste:

Se realizó ajustes de los modelos predictivos según sea necesario para mejorar la precisión de las predicciones.

PRUEBAS ESTADISTICAS - ANALISIS PREDICTIVO

Prueba de Hipótesis para la Media:

Supongamos que queremos probar si hay una diferencia significativa en los gastos promedio entre dos trimestres específicos (por ejemplo, el primer trimestre del año pasado y el primer trimestre de este año).

La hipótesis nula (H0) sería que no hay diferencia en los gastos promedio entre los dos trimestres, mientras que la hipótesis alternativa (H1) sería que hay una diferencia significativa.

Podríamos utilizar una prueba t de Student para comparar las medias de los dos grupos de datos y determinar si la diferencia observada es estadísticamente significativa.

Prueba de Correlación:

Si sospechamos que hay una relación entre los gastos y otra variable (por ejemplo, el número de clientes atendidos en cada trimestre), podríamos utilizar una prueba de correlación para evaluar la fuerza y la dirección de esta relación.

Podríamos calcular el coeficiente de correlación de Pearson entre los dos conjuntos de datos y determinar si la correlación observada es estadísticamente significativa.

Aquí tienes un ejemplo de cómo se llevaría a cabo una prueba de hipótesis para la media de dos trimestres específicos:

Supongamos que tenemos los siguientes datos de gastos por trimestre:

Primer Trimestre del Año Pasado: [1000, 1200, 900, 1100, 950]

Primer Trimestre de Este Año: [1300, 1400, 1250, 1350, 1200]

Para realizar la prueba de hipótesis, seguiríamos estos pasos:

Formulación de Hipótesis:

H0: No hay diferencia en los gastos promedio entre los dos trimestres.

H1: Hay una diferencia en los gastos promedio entre los dos trimestres.

Cálculo de Estadístico de Prueba:

Utilizaríamos la fórmula para la prueba t de Student para calcular el valor del estadístico de prueba.

**Determinación del Valor Crítico o p-valor**

Utilizaríamos la distribución t de Student para encontrar el valor crítico correspondiente para nuestro nivel de significancia deseado (por ejemplo, 0.05) o calcularíamos el p-valor asociado al estadístico de prueba.

**TOMA DE DECISIÓN**

Si el valor del estadístico de prueba es mayor que el valor crítico o el p-valor es menor que el nivel de significancia, rechazamos la hipótesis nula y concluimos que hay una diferencia significativa en los gastos promedio entre los dos trimestres