**UNIVERSIDAD POLITÉCNICA DE MADRID**

MÁSTER UNIVERSITARIO EN INGENIERÍA INFORMÁTICA

Final Assignment #1

Project Planning

Autores:

* Pablo Fernández Díaz

(pablo.fernandez.diaz@alumnos.upm.es)

* Roberto García Álvarez

(r.galvarez@alumnos.upm.es)

Índice

[1 BACKGROUND 2](#_Toc531618891)

[1.1 OBJETIVO & ESTADO DE OPORTUNIDAD 2](#_Toc531618892)

[1.2 MEJORAS DEL SERVICIO 2](#_Toc531618893)

[1.3 CASO DE NEGOCIO 2](#_Toc531618894)

[2 VISION DEL PROYECTO Y ALCANCE 3](#_Toc531618895)

[2.1 CUESTIONES DE INVESTIGACIÓN 3](#_Toc531618896)

[2.2 ALCANCE & ÉXITO EN LOS RESULTADOS 3](#_Toc531618897)

[2.3 MEDIDAS DE ÉXITO / ROI 4](#_Toc531618898)

[2.4 ARQUITECTURA 4](#_Toc531618899)

[2.5 COMUNICACIÓN 5](#_Toc531618900)

[3 PLAN DEL PROYECTO 5](#_Toc531618901)

[3.1 ROLES & RESPONSABILIDADES 5](#_Toc531618902)

[3.2 FASES DEL PROYECTO 6](#_Toc531618903)

[3.3 ASPECTOS ECONÓMICOS Y PRESUPUESTO 7](#_Toc531618904)

[3.4 RIESGOS 9](#_Toc531618905)

# BACKGROUND

## OBJETIVO & ESTADO DE OPORTUNIDAD

El principal objetivo del **proyecto** es ofrecer al cliente un **incremento en las ventas** de los diferentes productos vendidos en el establecimiento. Durante todo el plan del proyecto se presentarán y defenderán las diferentes razones por las que la implementación del proyecto **producirá un beneficio** muy positivo para el cliente.

Para llevar a cabo el modelo de datos desarrollado se han planteado el problema desde un el campo de la **Analítica Asociativa**, ya que el principal objetivo es encontrar que **productos** se venden en más ocasiones juntos. Además, mediante el análisis de otras variables, se presentarán las mejores ofertas para incrementar al máximo los ingresos del cliente.

## MEJORAS DEL SERVICIO

El principal objetivo del proyecto es **conocer mejor los gustos de los compradores** basando las conclusiones finales en respuestas analíticas.

Obviamente, durante los diferentes días de la semana, y dividiendo cada día en los diferentes rangos horarios, podemos definir y describir diferentes preferencias en los productos vendidos y relaciones entre los productos que se venden juntos.

Por lo tanto, el principal cambio de servicio que los trabajadores tendrán que implementar es **brindar las diferentes ofertas a los compradores**. Estas ofertas deben ser programadas y desarrolladas por los responsables de la tienda en función de los resultados del análisis de los datos.

## CASO DE NEGOCIO

El objetivo principal de este proyecto es **mejorar los ingresos de una panadería** mediante la comprensión de las necesidades del cliente y el desarrollo de ofertas y distribución de productos en la tienda para obtener un mayor beneficio.

No hay nada que no funcione correctamente o que esté "roto" pero, obviamente, el rendimiento siempre puede mejorar.

Este proyecto es urgente ya que cuanto antes se encuentren respuestas a nuestras preguntas antes dejará de perder dinero el cliente. Así que, obviamente, las principales **consecuencias** de no implementar esta solución son **económicas**.

# VISION DEL PROYECTO Y ALCANCE

## CUESTIONES DE INVESTIGACIÓN

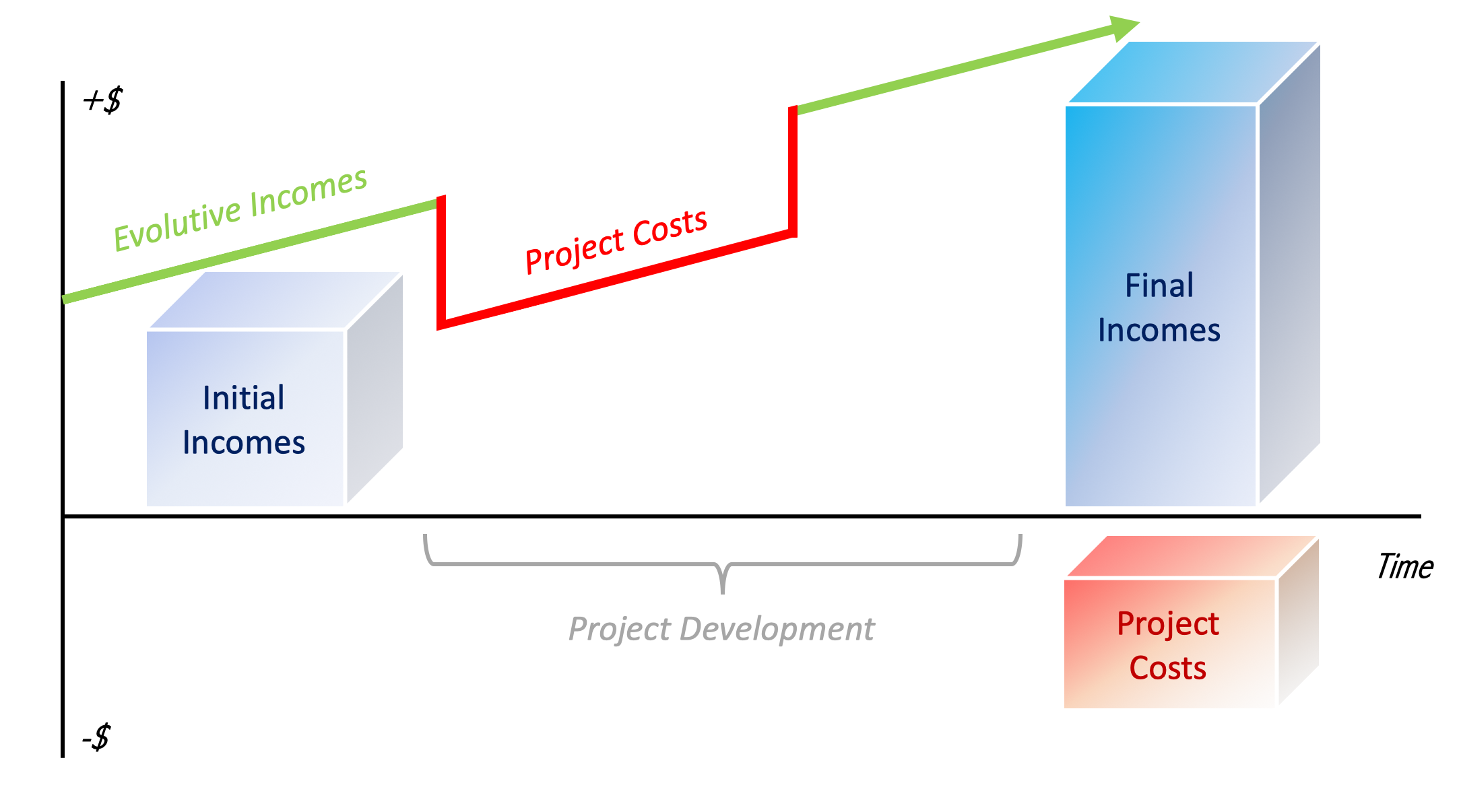
Antes de comenzar a desarrollar un proyecto de análisis de datos es muy importante realizar una investigación previa y analítica para averiguar cuáles podrían ser las preguntas que se pueden responder como resultado de nuestro modelo de datos:

* ¿Qué tipo de ofertas debemos hacer los fines de semana?
* ¿Deben ser ofertas diferentes según el momento del día?
* ¿Las ventas dependen de la ubicación?
* ¿Las ventas dependen del número de productos comprados?
* ¿Qué tipo de productos debemos vender juntos?

Para responder a estas preguntas utilizaremos el algoritmo A priori para poder encontrar reglas del estilo *Producto A and Producto B then Producto C*, siendo el Producto A, por ejemplo, tostada, el producto B café y como resultado un menú que sea café con tostada, Producto C. La finalidad es que estas reglas tengan una alta confianza y soporte para los datos que estamos analizando en nuestro intervalo de tiempo.

## ALCANCE & ÉXITO EN LOS RESULTADOS

Durante el presente plan de proyecto se presentarán los resultados y plazos para la **recuperación** del capital inicial de la **inversión total**. Entendiéndose, recuperación de la inversión como la obtención de beneficios, teniendo en cuenta los gastos del proyecto:



Este proyecto está enfocado en mejorar la experiencia del comprador y aumentar los ingresos de la panadería.

## MEDIDAS DE ÉXITO / ROI

Para justificar la inversión en el proyecto del modelo de datos, a continuación, se detallarán las estimaciones que recogen las **posibles mejoras en el negocio al aplicar el modelo**.

Según los datos iniciales prestados por el cliente en el periodo medido se han llevado a cabo un total de 7097 transacciones (**21293** artículos vendidos totales). Suponiendo que el precio medio de cada artículo equivale a 2 € y que el beneficio por artículo equivale al **40%,** en el periodo de datos medidos (6 meses) se habría obtenido un beneficio medio estimado de 17.034,40 €, por lo tanto, de media al mes se obtienen unos beneficios estimados de **2.839,07 €.**

Actualmente el negocio vende diversos productos, pero el objetivo de este proyecto es averiguar que conjunto de los mismos suele venderse juntos (basándose en reglas asociativas apoyadas por soporte elevado). Por lo tanto, se le propondrá al cliente **10 reglas** con mayor soporte representando los resultados del modelo. Dichas reglas podrán ser usadas en un futuro con el objetivo de crear ofertas y visibilizar los productos mas potenciales en ciertas horas y situaciones.

El objetivo, por lo tanto, del proyecto es averiguar cuales son los productos que se venderán juntos potencialmente. Es decir, tras la aplicación del modelo se pretenderá que las ventas aumenten, de forma que si inicialmente existían transacciones de un único artículo en ciertas horas en las que la venta de dos artículos conjuntos era elevada, se creen, a posteriori, ofertas que permitan que esas transacciones de artículos únicos se conviertan en transacciones de múltiples productos.

Suponiendo que en el 50% de los casos, tras aplicar el modelo, se consigan convertir las transacciones únicas supondría un aumento exponencial en los beneficios de la panadería. Partiendo de que la media de los artículos vendidos al mes son 3.548, esto supondría un **aumento a 5.323 artículos vendidos al mes**. Suponiendo que la media del precio de los productos es de 2€ y que el negocio obtiene un 40% de ganancia, en total, y tras aplicar el modelo, **podría aumentar sus beneficios 49.9%**, aumentando los ingresos a una **media mensual estimada de 4.258,40€.**

## ARQUITECTURA

**INFRAESTRUCTURA:**

El objetivo del proyecto es que sea reutilizable, por tanto, se realizara en la nube, teniéndolo organizado en plataformas como Github, con la finalidad de que se lleve a cabo un trabajo cooperativo y al que puedan acceder todos los desarrolladores implicados.

**HERRAMIENTAS:**

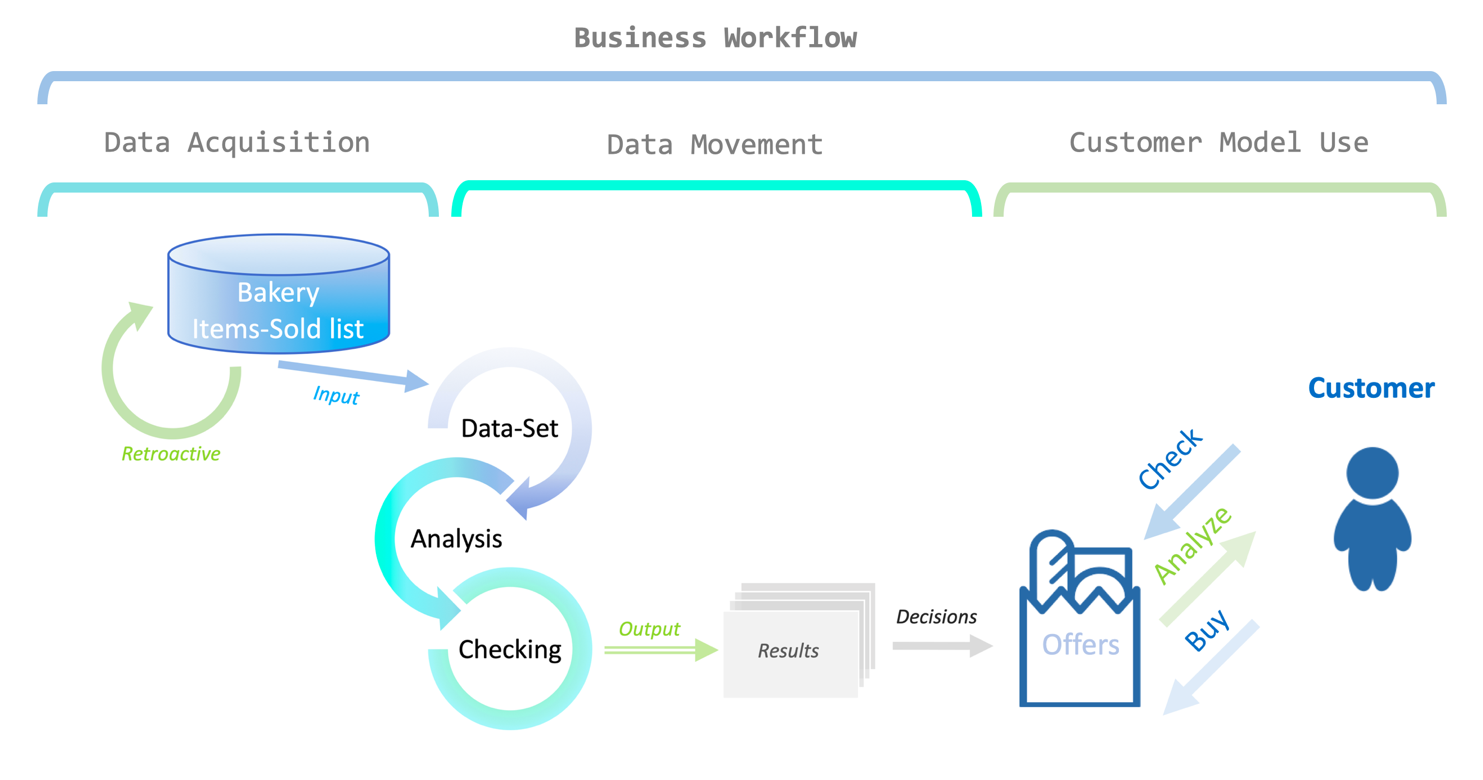
Para el desarrollo principal del proyecto decidimos usar R como lenguaje de programación base. R es un lenguaje y un entorno para la computación estadística y los gráficos. Es un proyecto GNU que es similar al lenguaje y entorno S que se desarrolló en *Bell Laboratories*. R puede ser considerado como una implementación diferente de S. Hay algunas diferencias importantes, pero mucho del código escrito para S se ejecuta sin alteraciones bajo R.

R proporciona una amplia variedad de técnicas estadísticas (modelos lineales y no lineales, pruebas estadísticas clásicas, análisis de series temporales, clasificación, agrupación, ...) y técnicas gráficas, y es altamente extensible. El lenguaje S es a menudo el vehículo elegido para la investigación en metodología estadística, y R proporciona una ruta de código abierto para la participación en esa actividad.

Centrándonos en el entorno con el que trabajaremos, decidimos utilizar RStudio, que es un entorno de desarrollo integrado (IDE) libre y de código abierto para R, un lenguaje de programación para computación estadística y gráficos.

**FLUJO DE TRABAJO:**

El flujo de trabajo entre la entidad desarrolladora y el cliente se fundamenta en la entrega e implementación del modelo y sus resultados. A partir de ahí, se buscará la obtención de beneficios ante la implementación de ofertas que **incrementen los beneficios** del comercio. De esta forma, el comprador final solo llevará a cabo una interacción con el negocio en la cual compruebe y valide las ofertas que se propondrán.

****

## COMUNICACIÓN

La comunicación entre el desarrollo del proyecto y el cliente debe ser frecuente y fluida. El calendario principal de las reuniones de comunicación, por lo menos, tiene que estar compuesto por:

1º **Reunión inicial**: Necesario para aclarar el objetivo principal del proyecto y las preferencias y prioridades del cliente.

2º **Reunión intermedia:** A mitad del proyecto se realizará una reunión intermedia para dar un feedback al cliente.

3º **Reunión final**: la reunión final tiene como objetivo principal presentar al cliente los resultados del desarrollo del proyecto y el precio y producto final.

Además, es muy importante proporcionar y explicar al cliente la documentación principal.

En el equipo de desarrollo, la persona que va a estar en contacto con el cliente será el **jefe de proyecto.** Y estará en contacto con el **propietario** y también con la persona que tiene la **máxima responsabilidad** en la panadería.

# PLAN DEL PROYECTO

## ROLES & RESPONSABILIDADES

|  |  |  |
| --- | --- | --- |
| Role (*id*) | Responsabilidad | Departamento |
| Jefe de Proyecto (*pm*) | Gestionar y controlar todas las responsabilidades y tareas.  Comunicarse con el cliente.  Controlar los departamentos (mantenerse en contacto con los gerentes). | Management |
| Experto de gestión de datos(*ds*) | Realizar las diferentes tareas de análisis de datos necesarias durante el desarrollo del proyecto de análisis de datos. | Data |

Tabla 3: Roles y responsabilidades

El siguiente diagrama muestra los rangos de responsabilidades de los diferentes roles del proyecto, siendo el Gerente de Proyecto el trabajador con mayor responsabilidad:

Data Department

Management Department

El gráfico anterior muestra el esquema de los roles de los trabajadores en el equipo. El Jefe de Proyecto tiene la mayor responsabilidad y la principal tarea de controlar y coordinar todos los procesos y tareas que se llevarán a cabo durante el desarrollo del proyecto. Además, es el máximo responsable y será el representante de la parte desarrolladora durante las diferentes reuniones e intercambio de información. El Jefe de Gestión de Datos es la persona que tiene que coordinar todas las tareas que realiza el Departamento de Datos. Dichas tareas se llevarán a cabo principalmente por parte de los expertos en datos: 2 Expertos en Gestión (que desarrollarán e implementarán el modelo de datos) y un experto en Administración de Datos (que preparará los datos antes de empezar con el desarrollo y además gestionará la tarea final de documentación del proyecto).

## FASES DEL PROYECTO

Al principio del proyecto el jefe de proyecto se reunirá con el cliente para verificar los datos que le otorguen. Posteriormente ya se reunirán el jefe de proyecto y el experto de datos para llevar a cabo el desarrollo del proyecto. La estimación de la duración del **desarrollo del proyecto** es de aproximadamente 4 días. La siguiente tabla muestra la distribución de los diferentes plazos y fases del proyecto:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Inicio | Fin | Fase del Proyecto | Trabajadores involucrados | Hito/Meta | Prioridad |
| 1-Dic-8:00 | 1-Dic-14:00 | Análisis de los datos & Procesado [+ tests] | *ds* | 1 | 2 |
| 1-Dic- 14:00 | 3-Dic-10:00 | Desarrollo del Modelo de Datos [+ tests] | *ds* | 2 | 5 |
| 3-Dic-10:00 | 3-Dic-16:00 | Validación del modelo | *Pm, ds* | 3 | 4 |
| 3-Dic-16:00 | 4-Dic-14:00 | Análisis de los Resultados | *Pm, ds* | 4 | 4 |
| 4-Dic-12:00 | 4-Dic-18:00 | Documentación | *pm, ds* | 4 | 3 |

Tabla 4: Fases del proyecto

Las dos primeras fases del proyecto poseen una pequeña **etapa de tests**. Dichas **pruebas automáticas** buscan verificar que el modelo que se está obteniendo no posee errores de estructura, significado y desarrollo.

Las principales fases del proyecto son:

- **Análisis y procesamiento de datos**: análisis los datos de origen y limpieza/pre-procesado (mejorando los datos eliminando el ruido y los valores perdidos).

- **Desarrollo del Modelo de Datos**: desarrollar un modelo para obtener las mejores ofertas en función de los datos de entrada.

- **Validación del Modelo**: verificación de los resultados obtenidos. Se pretende asegurar que el modelo desarrollado cumple los requisitos exigidos por el cliente.

- **Análisis de resultados**: analizar los resultados y conclusiones de los resultados del modelo de datos.

- **Documentación**: documentar los resultados para explicar al cliente los resultados del proyecto.

En lo referente con la con el apartado de **Prioridad** hemos dividido el desarrollo del proyecto en 5 niveles de prioridad. Siendo 5 el nivel más crítico y 1 el menor. Todas las fases del proyecto son importantes y deben ser **completadas con éxito**. Pero la prioridad de cada fase podría afectar a los plazos, si el gestor del proyecto lo considera positivo para el desarrollo final del proyecto.

## ASPECTOS ECONÓMICOS Y PRESUPUESTO

La inversión que debe llevar a cabo el cliente se fundamentará en el coste total del desarrollo del proyecto. Dicho coste se divide en dos partes; en primer lugar, el sueldo de los trabajadores que desarrollaran el proyecto, y, en segundo lugar, el coste asociado a dichos trabajadores, lo cual equivale al **20%** del coste inicial.

Los diferentes trabajadores representan diferentes roles y cada rol tiene un salario específico por hora:

- Jefe del Proyecto 30 € per hora \* (6 + 8 + 6) horas

- Experto de datos 20 € per hora \* (8 + 12 + 6 + 8 + 6) horas

Cada trabajador llevará a cabo diferentes tareas, por lo tanto, el coste asociado al sueldo del mismo dependerá de la cantidad de esfuerzo individual destinado.

|  |  |  |  |
| --- | --- | --- | --- |
| **Tareas** | **Horas** | **Trabajadores Involucrados** | **Esfuerzo Total** |
| Análisis de los datos & Procesado | 8 horas | 1 | 8 h. |
| Desarrollo del Modelo de Datos | 12 horas | 1 | 12h. |
| Validación del modelo | 6 horas | 2 | 12 h. |
| Análisis de los Resultados | 8 horas | 2 | 16 h. |
| Documentación | 6 horas | 2 | 12 h. |

La siguiente tabla muestra la distribución en horas del esfuerzo total para las tareas:

Tabla 5: Esfuerzo Total para la realización de las tareas

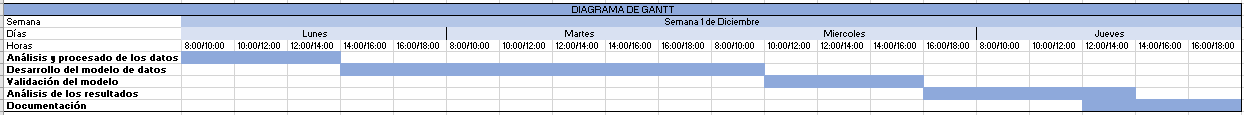
De acuerdo con los salarios de nuestros trabajadores y el esfuerzo total por tareas que realizarán, podemos considerar que el precio del trabajo de desarrollo del proyecto es:

|  |  |  |  |
| --- | --- | --- | --- |
| Fase | Trabajadores | Horas totales | Coste |
| Análisis de los datos & Procesado | *ds* | 8 h. | 160 € |
| Desarrollo del Modelo de Datos | *ds* | 12 h. | 240 € |
| Validación del modelo | *pm, ds* | 12 h. | 300 € |
| Análisis de los Resultados | *pm, ds* | 16 h. | 400 € |
| Documentación | *dm, ds* | 12 h. | 300 € |
|  |  | Salarios: | 1400 € |
|  |  | **Total:** | **1680 €** |

Tabla 6: Costes de los trabajadores del proyecto

En el coste total también se han tenido en cuenta otros tipos de gastos como puede ser los que vaya a tener el cliente en ese periodo, teniendo en cuenta los gastos de luz, agua... Para estos gastos hemos concluido que con un 20 % más se cubrían.

El siguiente diagrama de Gantt muestra la distribución del desarrollo de la tarea durante el período de tiempo (1 de diciembre a 4 de diciembre). Se consideran estos cuatro días como días laborables.



## RIESGOS

Para poder tener controlado la mayoría de los inconvenientes que nos podemos encontrar a la hora de desarrollar el proyecto y los cuales pueden demorar la finalización de este, hemos previsto las siguientes circunstancias que son las que más problemas nos darían.

- Que el cliente no pueda asistir a las reuniones planteadas o no nos pueda facilitar la información que le pidamos según vamos avanzando en el proyecto, para ello, si no puede asistir en persona se le llamará o se concertara una video llamada para poder avanzar en el proyecto e ir proporcionándole el *feedback* necesario.

- Que no se cumplan los plazos por parte de la entidad desarrolladora, para ello se realizara una penalización sobre el total del presupuesto en función de los días que se haya retrasado la entrega.

- Que se estropeen las máquinas con las que se estén trabajando. Si esto sucediese se dispondrá de nuevas máquinas para poder finalizar el proyecto.