**Project Proposal**

[**1.**](#_heading=h.5h7q3342enm4) **Current Situation-Background 2**

[**2.**](#_heading=h.p366gtslhqv8) **Proposal Objectives 2**

[**3.**](#_heading=h.6aw7hgmjbnk8) **Reach 3**

[**4.**](#_heading=h.wf6hudgiksa1) **KPI list to analyze 3**

[**5.**](#_heading=h.9yt9wqvsset8) **Proposal solution 4**

[**6.**](#_heading=h.bcir4aeiq37q) **Technological Stack 4**

[Python](#_heading=h.fbmfl5x1l4jb) 4

[AWS](#_heading=h.8y64r38tsxca) 4

[GitHub](#_heading=h.dim3uepzcgsv) 5

[**7.**](#_heading=h.cbiwnmmoqbr) **Work Methodology 6**

[**8.**](#_heading=h.yroahh6chw0h) **Detailed design – deliverables 7**

Week [1.-Project Proposal.](#_heading=h.i8cwyvpkgu4w) 7

Week [2.-Data Engineering.](#_heading=h.q1idyrtbzcu3) 7

Week [3.-Data Analytics y ML](#_heading=h.97zzhmiso8mq) 7

Week [4.-Final Presentation](#_heading=h.wiz494k072mk) 7

[**9.**](#_heading=h.g2vxnnwaeapc) **Team Roles and responsibilities 8**

1. Current Situation-Background

Olist was born as Solutions in 2007 with a different business idea than the one they currently manage, it was not until 2015 that they acquired the model they manage today.The objective of Olist, a Brazilian company, is to connect small companies with larger markets. In order to sell products to a larger number of customers.

Olist's solution is to help companies increase online sales for small and medium-sized retail companies. In addition, to make operations more efficient through integrated tools in the sales channels. In order to eliminate the fragmented experience that the market imposes on those looking for ways to digitize business.

Olist has the goal to continue growing in the Latin American region, for which it is committed to request consulting from an accredited firm specialized in Data Analytics and Data Science, which provides the analysis of information that allows taking the right steps to cover the annual commitment to increase their income.

1. Proposal Objectives

OList aims for annual sales growth; the actions to be taken will be based on the results of the data analysis. To achieve this goal, the following objectives were set:

* Increase our client's sales through:
  + Reduce the average delivery time of orders
  + Provide strategic recommendations to sellers
  + Expand the business model in Latin America

1. Reach

This proposal has the next scope:

* Deliver data analysis according to established objectives.
* Prepare an incremental load and model run for delivery of updated information.
* Development of Machine Learning models that allow:
  + Evaluate sentiment analysis
  + Develop the system of strategic recommendations to sellers

This Proposal does NOT contemplate::

* Installation of recommended processes.
  + Construction of logistics centers
  + Execution of the expansion in Mexico
* Development of new KPIs
* Any analysis that is not specified in this proposal

1. KPI list to analyze

Based on our experience in similar situations with previous clients, the average improvement when applying these proposals is around 10% (subject to analysis in the next stage of the project)

* To improve CRR (Customer Retention Rate)

Evaluating the rate of repeat purchases.

* To Increase AOV (Average Order Value)

Applying related purchase recommendations.

* To reduce delivery times between Carrier and Seller (CSDT)

Recommending the creation of logistics centers

* Improving CSAT (Customers Satisfactions) Score

Performing a Sentiment Analysis

* To increase Conversion Rate (CR)

Recommending Marketing campaigns based on the information collected

* To increase Olist penetrations in Latin America

Performing metrics related to the demographics and economy of each country.

1. Proposal solution

Our business proposal is mainly based on:

- Deliver a model of a product recommendation system for suppliers.

* We understand that it is important for Olist to provide information to sellers that will allow them to increase their sales, which is why we considered establishing a recommendation system for products that were also sold when the sale was made. Information will be taken from the orders, based on the preferences or previous behavior of the users.

- Analysis of available data that allows entering new markets in Latin America.

* Through an analysis, it was detected that Mexico is the second country, after Brazil, with the highest growth in electronic commerce within the region (data published by AMVO). Mexico presents a greater facility than other areas due to its high demographic concentration centralized in the Federal District. Finally, this country represents the 2nd largest economy in Latin America (source CEPAL).

- Analysis of products with high turnover that allows the seller to offer the service of Olist Logistics Centers.

* According to metrics of different E-commerce, having their own storage and distribution centers decreases delivery times, which results in an average increase in sales of 63% (source AMAZON).

1. Technological Stack

Python

To develop the EDA, **Python** was used along with its **Pandas** and **Matplotlib** libraries.

AWS

We decided to go with the AWS service because it has these advantages over the rest:

* **Scalability:** AWS allows for easy scaling of resources according to the needs of the application, whether increasing or decreasing processing, storage and bandwidth capacity.
* **Security:** AWS has a wide variety of security options and services, including authentication, compliance, data protection and policy enforcement.
* **Broader range of services:** AWS offers more services than others which makes it more versatile and able to handle a wider range of workloads.
* **More mature ecosystem:** AWS has been around longer than its competitors and as a result has a more mature ecosystem with more third-party tools and services available.

The services that were decided to use from AWS are the following:

* **IAM (Identity and Access Management)** is an AWS service that allows administrators to control who has access to the resources of the AWS account and under what circumstances.
* **S3 (Simple Storage Service)** is an AWS cloud storage service that allows to store and retrieve data through a standard web interface.
* **Glue** is an AWS service that allows to create and manage a data catalog to facilitate data integration between different AWS services. Glue also provides an extraction, transformation, and loading (ETL) engine to help clean, transform and move data.
* **Lambda** is an AWS service that allows users to run code without having to provision or manage servers. With Lambda, can create functions that are automatically triggered when an event occurs (such as a data upload to S3 or an HTTP request) and automatically scale according to demand.
* **Athena** is an AWS service that allows SQL queries on data stored in S3 without having to provision or manage infrastructure. With Athena, can analyze data quickly and efficiently.
* **QuickSight** is an AWS service that allows it to create and share data visualizations and reports quickly and easily.

GitHub

Give us a back-up service with all the historic work

1. Work Methodology

It is committed to agile methodologies where SCRUM is chosen as the framework for the project.

The following work scheme is defined

**PO-Product Owner**- [Francisco Negrete](mailto:fnegrete@soyhenry.com)

**SM- Scrum Master**- Gonzalo Posse

**ST-Scrum Team:**

* Damián Bush
* Gisela Medina
* Martin Menéndez
* Mauro Cadme
* Miguel Angel Salas

Dailys are coordinated, reported to the Scrum Master to inform the daily progress, the incidents found and the tasks committed for the next day, by each team member.

Each week a sprint Goal is defined, which is delivered to the PO on Fridays, There we get feedback and agree upon the sprint Goal for the following week.

Take into account the good practice of not accepting changes that jeopardize the delivery of the sprint Goal that the team has in process.

The team is in charge of carrying out the tasks defined in the to Do's list by each weekly Sprint Planning. Our commitment is::

* Give the best quality to the deliverable.
* Adapt our plan each day towards the objective.
* Be responsible and professional in each of the tasks.
* When each one finishes their part, support other team members.

1. Detailed design – deliverables

Deliverables List. The following documentation will be delivered, according to each week.

Week 1.-Project Proposal.

* Detailed document on PDF
* PowerPoint presentation to thel PO
* Gantt diagram on PDF.
* Readme of Preliminary Data Quality Analysis

Week 2.-Data Engineering.

* ER model on pdf
* Pipelines to feed DW
* Data dictionary
* Workflow document detailing used technologies.
* Chosen stack and support of use.
* Data validation report.
* Readme.

Week 3.-Data Analytics y ML

* Dashboard
* Machine Learning model.
* Gantt diagram on PDF.
* Readme

Week 4.-Final Presentation

* GitHub Repository with all data and results
* Presentation Video

1. Team Roles and responsibilities

Define Scrum Team, all of them need to support the team; detailed roles:

* Damián Buch (CTO) - Data Engineer
* Gisela Medina (CEO) - Data Analyst
* Martin Menéndez(LDA) - Data Analyst
* Mauro Cadme(CDO) - Machine Learning
* Miguel Angel Salas (CPO) - Data Scientist

The entire development team contributes at each stage and is responsible for developing the product, self-organizing and self-managing. Whenever you finish a task, you mark it in Trello to let the group know it's done.