

## # Technical Questions

**First name:** *Fermín*

**Last names:** *Leal Payá*

This document is to resolve the technical questions proposed in the Data engineer test.

# Questions

**1.How long did you spend on the coding test? What would you add to your solutions if you had more time? If you didn't spend much time on the coding test then use this as an opportunity to explain what you would add.**

To answer this question, we will present 3 tables, one for each task which will specify the time spent in each

### **Task\_1: Clean Data and export a final output**

Task	Time
Analysis of the problem	0h 30m
Development approach	0h 30m
Development	3h 15m
Total Time	4h 15m

If we had more time, we could have parameterized the function \*\*\*

cleandata \*\*\* that is inside the program. This would mean that we would have to do the same as we do in the delivered function (look for bugs and fix them) but automatically.

Pseudocode example of this automatic function:

```
def cleandata (df,typeofcolumn,column,arrayoftotalcolumns,typeofcolumns):  
    #check column  
    #for example is type is number  
    df[column+'ok']=df[column].astype(str).apply(numbers)  
    df[column+'name']=df[column].astype(str).apply(isstring)  
    df[column+'email']=df[column].astype(str).apply(isemail)  
    #Clean bad positions  
    df[column] = np.where(df[column+'ok']!='', df[column+'ok'], '')  
    for i in arrayoftotalcolumns:  
        #do automatic clean
```

## Task\_2: Write an app that can generate core business events

Task	Time
Analysis of the problem	0h 30m
Development approach	0h 30m
Development	2h 45m
Total Time	3h 45m

If we had had more time we would have dedicated it to the function ***data*** within the script that deals with generating the orders. The way it is done is very simple, we count from 1 to 5 and generate 4 accepted orders and one canceled with an if condition. This can be improved with a loop that has a range of 0 to 4, within the other loop that already exists in the ***data*** function. So this second loop with an if when it reaches 4 it generates an order canceled instead of accepted.

### Task\_3: Write a streaming app

Task	Time
Analysis of the problem	0h 10m
Development approach	0h 5m
Development	1h 00m
Total Time	1h 15m

This function has been written in python but when it comes to making a monitor of a service we like to use platforms like ***Nagios*** already predestined to these services. ***Nagios*** is a service monitor that we have normally used to monitor processes in amazon web services and we think it would have been very good for this exercise.

### 2.Why did you choose the language you used for the coding test?

We have used python because it is one of the programming languages that we dominate the most. Also in the field of data science and data engineer

python is in the lead along with R since it has very powerful data processing tools such as pandas, tensorflow, etc...

It is also very fast to develop and test

### **3.What was the most useful feature that was added to the latest version of your chosen language?**

We personally really like the python *scikit learn*, *pandas* and *numpy* libraries that are also constantly updated with improvements.

In addition, since over two years we have very powerful deep learning libraries such as *Tensorflow* that belongs to google. *Tensorflow* has been used these years in many places, even as a curious fact, in the dynamic faces of *Snapchat*.

### **How would you track down a performance issue in production? Have you ever had to do this?**

This question is very broad, we can personally talk about these problems within the field of databases.

Many databases such as Redshift have their own monitor which shows what the processes are running, and you can even save a daily log if you want. If you have a monitor which checks the daily log and in one of the processes runtime has upper than other days something is surely happening.

For example, if we are running a daily load, one of the files to be loaded may have come with more data than normal or simply the file is wrong.

## 5. Please describe yourself using JSON.

```
{  
  "name": "Fermín",  
  "lastname": "Leal",  
  "age": 24,  
  "nationality": "Spanish",  
  "livesin": "Novelda, Alicante",  
  "passions": [  
    "Programming",  
    "Data Science",  
    "Data Engineering",  
    "Automobile",  
    "Travelling"  
  ],  
  "interests": [  
    "Science",  
    "Investment Market",  
    "Cryptocurrencies",  
    "Bicycle routes",  
    "Cinema"  
  ]  
}
```