Loggi Forecasting

Buenos Aires | Team #4

Leandro Benetton, Carlos Cernello,Braulio Soares, Pablo Manzano

Business Problem

# Loggi’s mission is to connect Brazil, delivering anything to anyone as quickly as possible. Through technology (mobile, AI, automation, IoT), Loggi has created a next-generation logistics network and is, in an unprecedented way, positioned to unleash the growth of a new trade-in Brazil with a fast, cost-effective and reliable logistics solution.

# Loggi’s current network is composed of tens of thousands of partner drivers connecting customers to hundreds of small hubs distributed all across Brazil, responsible for local operations. We also have a couple of large cross-docking facilities connecting the small hubs through a large scale network of ground and air transportation, responsible for national operations. With this model, we want to be present in all Brazilian cities by the end of 2020.

New requirements for e-commerce customers that have emerged in recent years have brought greater complexity to the logistics process, making last-mile delivery activity one of the most expensive and least efficient steps in the parcel delivery supply chain.

At Loggi, the last-mile activity includes distribution centers, named as agencies, in the main cities of the Brazilian territory. Depending on the delivery volume, a city can have several agencies. The last-mile delivery itself is carried out through autonomous drivers (independent contractors), a fact that implies some specific challenges for efficiently allocating drivers to delivery routes.

Currently, Loggi’s agencies know the demand of last-mile deliveries for a given day, but have a reduced visibility on the supply (flow) of drivers during the hours of that day - what impacts its capacity to allocate drivers to itineraries. Besides, it is known that driver habits related to hour of the day (time slot), days leading up to and following holidays, rainy days, strong traffic days, etc. could affect the driver's motivation to accept a new itinerary. Analyzing these data could make it feasible to understand how much demand an agency is able to meet in a specific day hour and during the day. The visibility of the hourly capacity would allow agencies to best distribute and schedule the process for allocating drivers during the hours of the day.

The DS challenge is creating a computational tool to analyze the current flow of Loggi's registered drivers during a given day, for a specific agency, and predict the capacity of this agency to allocate drivers to itineraries for that day, in an hourly base. This tool is intended to support immediate operational actions, as well as aligning expectations about short-term allocation potential.

Business Impact

Delivery time directly impacts the platform's user experience (UX) and the ecosystem health from the operational point of view since it increments the utilization of the fleet. It could be valuable for Loggi to better understand which agencies are underperforming, where it needs to recruit more riders and which areas are being underserved, in order to improve its operations.

Aiming to provide an initial analysis with a set of data that can change operational profitability and even make the customer transparent with the value of the service and while maintaining loyalty.

Stable time for the delivery of the weather function , traffic , day of the week and time.

They are the identifiers that we will use together to reduce the business impact.

# Data ● **Our ​Itinerary** data set includes information about the itineraries performed by the driver, such as creation date, agency, time of start, time of end, place of start, place of end, total distance, count of packages, etc

●  **Our ​Driver Availability** ​data set includes information about history of driver availability and geographical locations when using Loggi’s app   
●  **Our ​Distribution** Center ​data set, with information about all Loggi’s operational facilities including geographical locations   
● **Traffic**​ data set   
● **Weather** ​data set   
● **Holidays​** data set

# Methods

**Visualizations**

One key component of understanding the patterns in Loggi is providing high-quality visualizations of his historical delivery. Here are some of the static and interactive visualizations we will provide:

* Heatmaps
  + Delivery time map to understand underserved areas and analyze the impact of distribution center allocation
  + Demand heatmap to understand demand location
  + Rider density heatmap to understand fleet size and coverage (by itself and in relation to the demand)
* Time series
  + Week-day demand seasonality
  + Intra-day peak demand hours
  + Intra-day delivery time variations (most likely spike during peak demand hours)
  + Understanding how the company’s demand, fleet have been behaving in the long term; possibly exponential growth.
* Box-plots
  + Understanding delivery time in different weather conditions
  + Understanding delivery time in hourly/weekday
* Visualization about the relationship between the day of the week and time of the order and the availability of the distributors.
* The weather vs the availability of delivery people and duration of the trip
* The weather vs the time of the order and qty of orders
* Traffic vs delivery availability and duration of the trip
* Correlation plot about different availability and different variables.

**Models**

To model the delivery time we propose to break it into its component intervals. As we understand it the total delivery time on a delivery lifecycle could be understood as the sum of the following:

* Acceptance time (AT): The time interval between when the delivery is requested, the dispatching algorithm assigns it to a rider (or riders) and the rider accepts it. This time could be impacted by the relation between the demand and the rider availability on any given place and time. It would be interesting to study demand seasonality (daily and weekly) to determine peak hours, and climate conditions, to see if there is a scarcity of riders during rainy days. The restrictions in the dispatching algorithm can also have an incidence in this time interval, so it would be prudent to ask the company for more information about it. For example: does it allow for multiple orders to be assigned and stacked to a single rider? What does it optimize (delivery time or cost)?
* Start Time (ST): The start time is the difference of when the rider accepts an order and it actually starts fulfilling it. This Could be due to multiple orders assigned to a single rider. It can be extracted from the data we have and it most likely be correlated with the relation between the demand and rider availability in the city.
* Start to pick up time (SP): This interval is the time difference between when the rider starts fulfilling the order and it arrives at the pick-up point. It would be interesting to study the relationship between this time interval and the distance between the start point of the rider and the pick up point, as well as the vehicle type of the rider.
* Pick-up time (PU): The pick up interval is the time the rider spends at the pick-up point. It could be useful to use a classification algorithm to determine if an order has a Loggi distribution center as the pick-up location or it is a courier order, where the pick-up location can be a restaurant or an office, etc. Also the demand rate could impact this time interval, at peak hours on food delivery and on underperforming distribution centers.
* Pick up to delivery time (PD): Similarly to the SP time, the pick-up to delivery time will most likely be correlated with the distance the rider has to travel, his vehicle type and the traffic conditions in the city at the time. The area covered by each of Loggi’s distribution centers could be easly mapped during the analysis of this stage.

As to the techniques used to model each of the stages of the delivery time; they will most likely be regression analysis for the most part, but we can apply any of the analysis covered on the course. Moreover, we will need to use a classification algorithm to determine which pick-up points are distribution centers and which are custom locations.

To conclude, we think that the best way to approach the study of Loggi’s delivery time is to break it up in its component intervals and to analyze what variables impact each one of those. We have Loggi’s historical proprietary data and public climate and traffic data to perform the study and we’re looking forward to better understanding what are the underlying drivers of incremental delivery time.

# Interface

The final front-end product will feature two landing pages: an **Analytics** page with visualizations of the historical data about delivery, and a **Forecasting**  page, where the results are summarized with a recommendation about each features impact in time of delivery for example: weather and traffic



# Milestones

In this section, we provide details on the milestones we intend to achieve in our project. In particular, we have outlined four different versions: we expect to finish Version 1 with 100% probability, Version 2 with ~70% probability, Version 3 with ~20% probability, and Version 4 with ~5% probability (if things go extremely well).

Version 1: Build simple dashboard (2 static plots, 2 interactive plots) to understand the historical data information on trip orders times and its correlation with 3 different variables (weather, traffic and drivers availability) and the 3 different business operations.

Version 2: Build prediction model (+ 1 static plot, 1 interactive plot) and determine what is optimal time to ship an order and the better UX experience for each type of services

Version 3: Build appropriate causal inference model to assess the strength of the causal relationship between the variables studies and the time to ship the order (include key assumptions in the dashboard)

Version 4: Use additional on line data to adjust the prediction model by each hour.

# Timeline

|  |  |  |
| --- | --- | --- |
| **Date** | **Deliverable** | **Details** |
| **Week 1** | Team formation  Environment setup |  |
| **Week 2** | One-page summary  Workflow setup | *This includes getting data access, e.g. NDAs.* |
| **Week 3** | Scoping document  Data access | *The scoping document is the skeleton of the Final Report.* |
| **Week 4** | Data cleaning  Initial data exploration | *Update Final Report to include EDA results.*  *Business requests and interviews as possible.* |
| **Week 5** | Continue data exploration | *Get initial code reviewed by TA as well.* |
| **Week 6** | Advanced data exploration  Initial modeling | *Begin front-end visualization.* |
| **Week 7** | Continue modeling  Application on cloud | *Update Final Report to include initial modeling results.* |
| **Week 8** | Front-end complete  Advanced modeling | *Update Final Report to include final modeling results.* |
| **Week 9** | Fine-tune modeling  Fine-tune application | *Write Conclusions section of Final Report.* |
| **Week 10** | Finalize presentation, report, and application |  |

# Concerns

The availability of data on time to the EDA process and the understanding of the business core.