

DoorDash Regression Project

Problem objective: Build a model from the DoorDash data to predict the estimated time taken for a delivery.

Variable definitions: The attached file `historical_data.csv` contains a subset of deliveries received at DoorDash in early 2015 in a subset of the cities. Each row in this file corresponds to one unique delivery. All money (dollar) values given in the data are in **cents** and all time duration values given are in **seconds**.

Time features

- **market_id:** A city/region in which DoorDash operates, e.g., Los Angeles, given in the data as an id
- **created_at:** Timestamp in UTC when the order was submitted by the consumer to DoorDash. (Note this timestamp is in UTC, but in case you need it, the actual timezone of the region was US/Pacific)
- **actual_delivery_time:** Timestamp in UTC when the order was delivered to the consumer

Store features

- **store_id:** an id representing the restaurant the order was submitted for
- **store_primary_category:** cuisine category of the restaurant, e.g., italian, asian
- **order_protocol:** a store can receive orders from DoorDash through many modes. This field represents an id denoting the protocol

Order features

- **total_items:** total number of items in the order
- **subtotal:** total value of the order submitted (in cents)
- **num_distinct_items:** number of distinct items included in the order
- **min_item_price:** price of the item with the least cost in the order (in cents)
- **max_item_price:** price of the item with the highest cost in the order (in cents)

Market features

DoorDash being a marketplace, we have information on the state of marketplace when the order is placed, that can be used to estimate delivery time. The following features are values at the time of **created_at** (order submission time):

- **total_onshift_dashers**: Number of available dashers who are within 10 miles of the store at the time of order creation
- **total_busy_dashers**: Subset of above **total_onshift_dashers** who are currently working on an order
- **total_outstanding_orders**: Number of orders within 10 miles of this order that are currently being processed.

Predictions from other models

We have predictions from other models for various stages of delivery process that we can use:

- **estimated_order_place_duration**: Estimated time for the restaurant to receive the order from DoorDash (in seconds)
- **estimated_store_to_consumer_driving_duration**: Estimated travel time between store and consumer (in seconds)