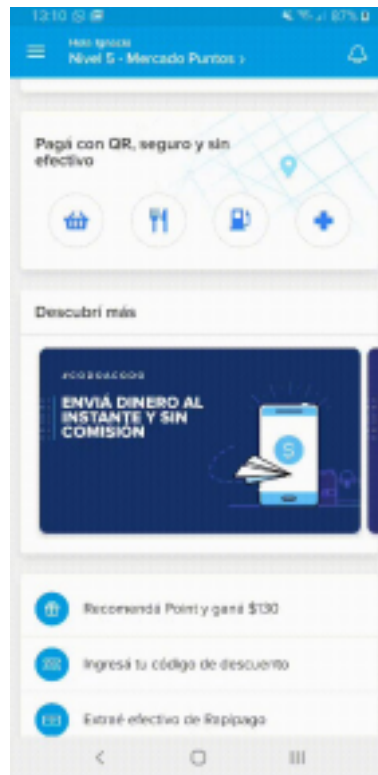


Description

En el contexto de MercadoPago, se quiere desarrollar un modelo de Machine Learning para predecir el orden de un conjunto de Propuestas de Valor (aka, Value Props) en el carrusel de la app llamado “Descubrí Más”.



La tarea consiste en construir utilizando Python y sus librerías un pipeline que tenga como input 3 fuentes de datos diferentes y genere como resultado un dataset listo para ser ingerido por el modelo.

Data Sources

- **Prints** (prints.json) - historial de 1 mes de value props que fueron mostradas a cada usuario, en formato json lines (un json por línea), por ej:

```
{"day": "2020-11-01", "event_data": {"position": 0, "value_prop": "cellphone_recharge"}, "user_id": 98702}
{"day": "2020-11-01", "event_data": {"position": 1, "value_prop": "prepaid"}, "user_id": 98702}
{"day": "2020-11-01", "event_data": {"position": 0, "value_prop": "prepaid"}, "user_id": 63252}
{"day": "2020-11-01", "event_data": {"position": 0, "value_prop": "cellphone_recharge"}, "user_id": 24728}
{"day": "2020-11-01", "event_data": {"position": 1, "value_prop": "link_cobro"}, "user_id": 24728}
{"day": "2020-11-01", "event_data": {"position": 2, "value_prop": "credits_consumer"}, "user_id": 24728}
{"day": "2020-11-01", "event_data": {"position": 3, "value_prop": "point"}, "user_id": 24728}
{"day": "2020-11-01", "event_data": {"position": 0, "value_prop": "point"}, "user_id": 25517}
{"day": "2020-11-01", "event_data": {"position": 1, "value_prop": "credits_consumer"}, "user_id": 25517}
{"day": "2020-11-01", "event_data": {"position": 2, "value_prop": "transport"}, "user_id": 25517}
```

- **Taps** (taps.json) - historial de 1 mes de value props que fueron clickeadas por un usuario, en formato json lines, por ej:

```
{
  "day": "2020-11-01",
  "event_data": {
    "position": 0,
    "value_prop": "cellphone_recharge",
    "user_id": 198702
  }
}, {
  "day": "2020-11-01",
  "event_data": {
    "position": 2,
    "value_prop": "point",
    "user_id": 3708
  }
}, {
  "day": "2020-11-01",
  "event_data": {
    "position": 3,
    "value_prop": "send_money",
    "user_id": 3708
  }
}, {
  "day": "2020-11-01",
  "event_data": {
    "position": 0,
    "value_prop": "transport",
    "user_id": 93963
  }
}, {
  "day": "2020-11-01",
  "event_data": {
    "position": 1,
    "value_prop": "cellphone_recharge",
    "user_id": 93963
  }
}, {
  "day": "2020-11-01",
  "event_data": {
    "position": 0,
    "value_prop": "link_cobro",
    "user_id": 94945
  }
}, {
  "day": "2020-11-01",
  "event_data": {
    "position": 1,
    "value_prop": "cellphone_recharge",
    "user_id": 94945
  }
}, {
  "day": "2020-11-01",
  "event_data": {
    "position": 2,
    "value_prop": "prepaid",
    "user_id": 89026
  }
}, {
  "day": "2020-11-01",
  "event_data": {
    "position": 0,
    "value_prop": "link_cobro",
    "user_id": 7616
  }
}, {
  "day": "2020-11-01",
  "event_data": {
    "position": 0,
    "value_prop": "link_cobro",
    "user_id": 63471
  }
}
```

- **Payments** (pays.csv) - historial de 1 mes de pagos realizados por los usuarios, en formato csv, por ej:

```
pay_date,total,user_id,value_prop
2020-11-01,7.043936396378688,35994,link_cobro
2020-11-01,37.362597487124475,79066,cellphone_recharge
2020-11-01,15.84302090723111,19321,cellphone_recharge
2020-11-01,26.26270194881884,19321,send_money
2020-11-01,35.35239084233181,38438,send_money
2020-11-01,20.951447929454982,85939,transport
2020-11-01,74.47751161115718,14372,prepaid
2020-11-01,31.516431799161893,14372,link_cobro
2020-11-01,83.76452022503628,65274,transport
```

Expected Result

El dataset a construir deberá contar con la siguiente información:

- **prints de la última semana**
- por cada print:
 - un campo que indique si se **hizo click o no**
 - **cantidad** de veces que el usuario **vio cada value prop** en las 3 semanas previas a ese print.
 - **cantidad** de veces que el usuario **clikeo cada value prop** en las 3 semanas previas a ese print.
 - **cantidad de pagos que el usuario realizó para cada value prop** en las 3 semanas previas a ese print.
 - **importes acumulados que el usuario gasto para cada value prop** en las 3 semanas previas a ese print.

Deliverables

- Código Python
- Un Doc con una breve descripción de las decisiones tomadas.