

Week 1 Intern Report

Chuqiao Feng

August 24, 2023

Overview

1 VOD Recommendation

- LDD-data basic analysis
- LDD-data income & exposure analysis
- LDD-data IP analysis

2 Recommendation System Survey

- Reinforcement Learning
- Large Language Model

LDD-data IP analysis

Dataset

- income data: /data_rr/vod_ldd_ecpm_daily_details_rr/
- vip data: /data_rr/vod_ldd_ecpm_vip_details_rr/
- media data:
/data_rr/ldd_media/child_tags_20220215.parquet

Notice: 这里仅使用近 14 天数据进行分析

baseCid	cid	pay_amt	sd_income	buy_times	duration_type	ip	ip_of_baseCid
JlRfwLGdYA1619589353988	...	28.00	16.8000	2.0	月卡	汽车世界	汽车世界
RYkFLhcgAh1608551836986	...	59.00	59.0000	3.0	季卡	NaN	宝宝巴士
ApQkswPzZM1654872253896	...	28.00	28.0000	2.0	月卡	汽车世界	NaN
ljBQVzHtwm1604037148198	...	1.00	1.0000	1.0	周卡	NaN	汽车世界
...
qJOSilczB1603373597958	...	28.00	28.0000	NaN	NaN	碰碰狐	碰碰狐
vyouLBZrvX1593583439097	...	1.00	1.0000	NaN	NaN	碰碰狐	碰碰狐
cpTksMugnu1653891844798	...	1.00	1.0000	NaN	NaN	宝宝巴士	NaN

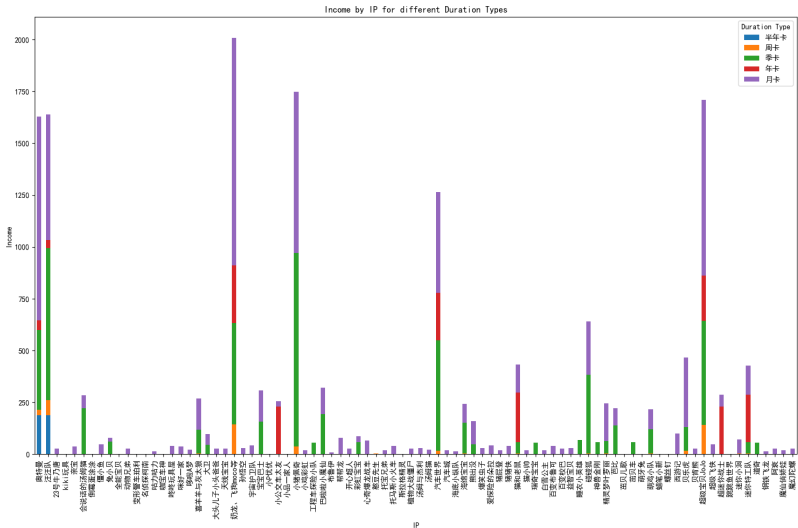
Table: Sample Merged Data

Duration Stats

Index	IP	IP of Base Cid	Duration Type	Pay Amount	SD Income	Buy Times
0	23 号牛乃唐	23 号牛乃唐	月卡	28.00	28.0000	5.0
1	kiki 玩具	kiki 玩具	周卡	1.00	1.0000	1.0
2	亲宝	亲宝	月卡	38.00	38.0000	14.0
3	会说话的汤姆猫	not-会说话的汤姆猫	周卡	1.00	0.5000	1.0
4	会说话的汤姆猫	not-会说话的汤姆猫	季卡	177.00	177.0000	6.0
...
206	道奇	not-道奇	季卡	55.00	55.0000	8.0
207	钢铁飞龙	not-钢铁飞龙	月卡	28.00	14.0000	1.0
208	阿衰	阿衰	月卡	28.00	28.0000	2.0
209	魔仙俏娇娃	not-魔仙俏娇娃	月卡	19.00	19.0000	1.0
210	魔幻陀螺	not-魔幻陀螺	月卡	28.00	28.0000	14.0

Table: 处理后的同系列 ip/非同系列 ip 统计数据 sample

Income by IP from different VIP duration types

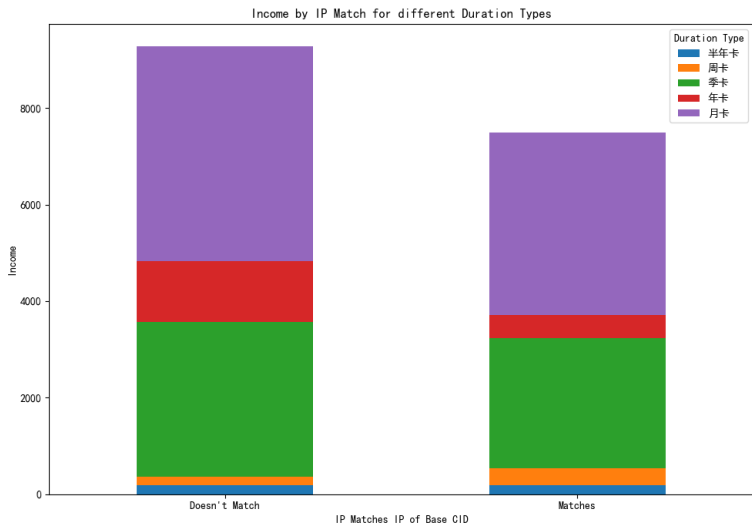


Income by IP from different VIP duration types

条形图显示了不同 vip 类型的所有 IP 收入, 从图表中可以得出:

- 某些 IP 在多种期限类型中产生大量收入
- 某些 IP 主要通过特定的持续时间类型赚取收入

Income by IP match for different VIP duration types



Income of IP match from different VIP duration types

条形图根据对于不同 vip 间类型下，推荐 cid 与 base cid 是否同系列 IP 来说明收入：

- 对于同类型 IP，所有期限类型的收入都会显着提高。
- 对于某些 vip 类型收入差异尤其明显，这意味着同系列 IP 在这些情况下可能具有特殊优势。

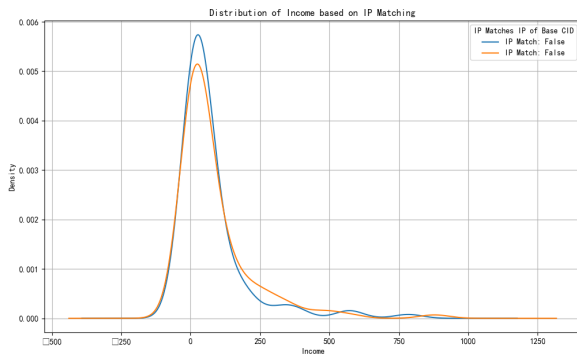
这表明，与非同系列的 IP 相比，与其 base cid 同系列的 IP 往往会在不同 vip 类型中产生更多收入。

- 按照 IP 是否同系列分类，对收入进行统计得到如下表格：

ip_match	count	mean	std	min	25%	50%	75%	max
0.0	101.0	74.224752	128.483013	0.17	6.6000	28.0	69.0	784.06
1.0	110.0	84.284545	136.442401	0.33	2.5625	28.0	89.0	878.41

Table: Match IP - Income discription

Distribution of Income based on IP matching



- 两个类别峰值均在较低收入值附近，低收入的情况较多。
- 同系列 IP 分布稍宽，比非同系列 income 分布范围更广。
- 分布相似，无较大差异。

Buy times information

- buy times: 检查是否同系列 IP 之间的购买次数是否存在显著差异。
- average income: 计算两个类别的每次购买的平均收入，了解到同系列 IP 倾向于进行更高价值的交易。

ip_match	total buy_times
False	1409.0
True	1529.0

total buy times

ip_match	avg_income
False	5.320582
True	6.063636

avg income per-purchase

Thank you!