

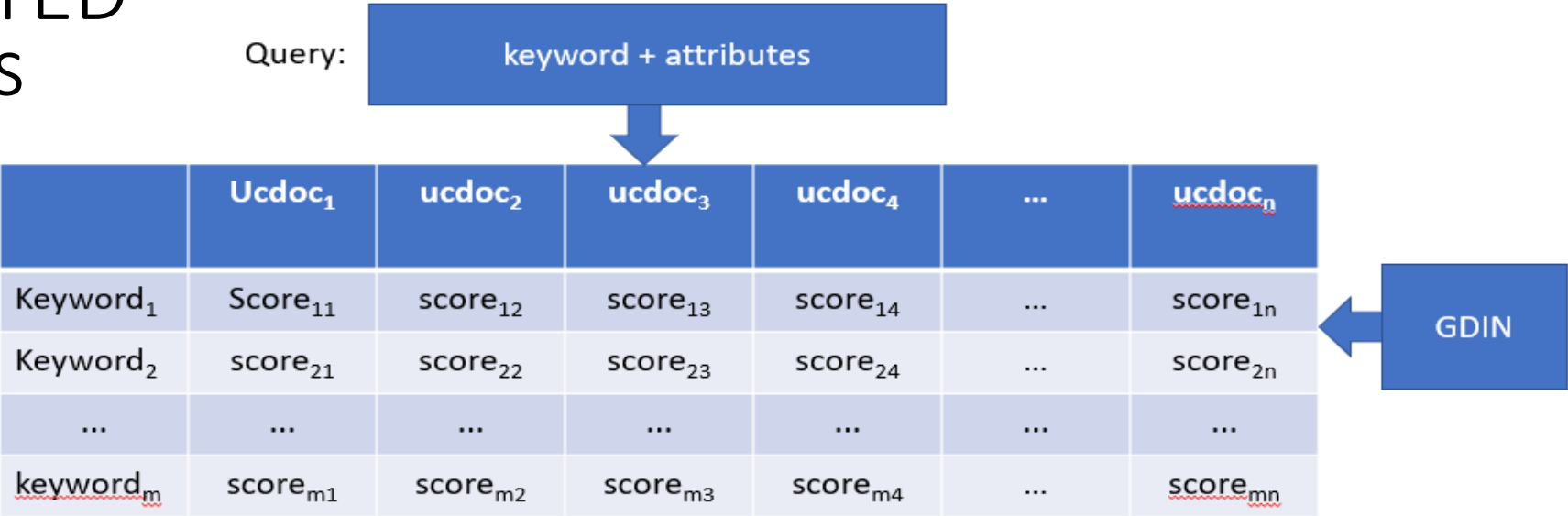
DIN Performance Forecasting

Main Menu

- Testing Data Preparation
 - Generate DIN predicted data on hive with DIN Predictor API
 - DIN prepared fact data on hive after DIN Pre-processing.
- Testing Processing Logic
 - DIN Model Recommendation VS Expert Experience VS Random Selection
- Testcases Design Rules
 - Based on the combination of specific attributes and creative keywords.
 - Three methods of selecting ucdocs: by din model, by expert experience, by random selection.
- Testing Result with Testcases Designed.

Testing Data Preparation: Generate DIN predicted data on hive with DIN Predictor API.

GDIN DATA
PREDICTED
UCDOCs



Output Table Design:

Ukey (as train ready)	attrs	kw1	kwn	Rank ₁ ucdoc
native, AI assistant, WIFI, 0, 5, 6	6 columns	0.08	0.02	Rank ₂ ucdoc
native, AI assistant, WIFI, 0, 5, 29	6 columns	0.07	0.03	Rank ₃ ucdoc
native, Huawei Video, 4G, 1, 6, 46	6 columns	0.07	0.01	Rank ₄ ucdoc
native, Huawei Music, WIFI, 0, 4, 1	6 columns	0.05	0.02	...
splash, Huawei Music, WIFI, 1, 3, 70	6 columns	0.08	0.05	Rank _n ucdoc

DIN Predictor API: input the keyword and ucdoc, output the score.

Input: a ucdoc with its history data of keywords and log counts, and the keyword.

e.g.
record = {"ucdoc": 1, "show_counts": ['25:3', '29:6,25:2', '29:1,25:2,14:2', '14:1,29:2,25:2', '29:1', '26:1,14:2,25:4', '14:1,25:3']}

Output: float numbers which present the ucdoc's scores for keywords.

e.g. [[0.0497351885], [0.0241885]]]

6 columns: gender, age, media, media_category, net_type, region_id.

Testing Data Preparation: DIN prepared fact data on hive after DIN Pre-processing.

GDIN DATA FACT UCDOCS

GDIN Train Ready Fact Data: GUCDOC

```
root
|-- region_id: integer (nullable = true)
|-- age: integer (nullable = true)
|-- gender: integer (nullable = true)
|-- net_type: string (nullable = true)
|-- media_category: string (nullable = true)
|-- media: string (nullable = true)
|-- uckey: string (nullable = true)
|-- interval_starting_time: array (nullable = true)
|   |-- element: integer (containsNull = true)
|-- keyword_indexes: array (nullable = true)
|   |-- element: string (containsNull = true)
|-- keyword_indexes_click_counts: array (nullable = true)
|   |-- element: string (containsNull = true)
|-- keyword_indexes_show_counts: array (nullable = true)
|   |-- element: string (containsNull = true)
|-- keywords: array (nullable = true)
|   |-- element: string (containsNull = true)
|-- keywords_click_counts: array (nullable = true)
|   |-- element: string (containsNull = true)
|-- keywords_show_counts: array (nullable = true)
|   |-- element: string (containsNull = true)
|-- uckey_index: integer (nullable = true)
|-- media_index: integer (nullable = true)
|-- media_category_index: integer (nullable = true)
|-- net_type_index: integer (nullable = true)
|-- gender_index: integer (nullable = true)
|-- age_index: integer (nullable = true)
|-- region_id_index: integer (nullable = true)
```

GDIN Logs Fact Data: GUCDOC Breakdown

```
root
|-- did: string (nullable = true)
|-- is_click: integer (nullable = true)
|-- action_time: string (nullable = true)
|-- keyword: string (nullable = true)
|-- keyword_index: integer (nullable = true)
|-- media: string (nullable = true)
|-- media_category: string (nullable = true)
|-- net_type: string (nullable = true)
|-- gender: integer (nullable = true)
|-- age: integer (nullable = true)
|-- adv_id: string (nullable = true)
|-- action_time_seconds: integer (nullable = true)
|-- interval_starting_time: integer (nullable = true)
|-- uckey: string (nullable = true)
|-- region_id: integer (nullable = true)
```

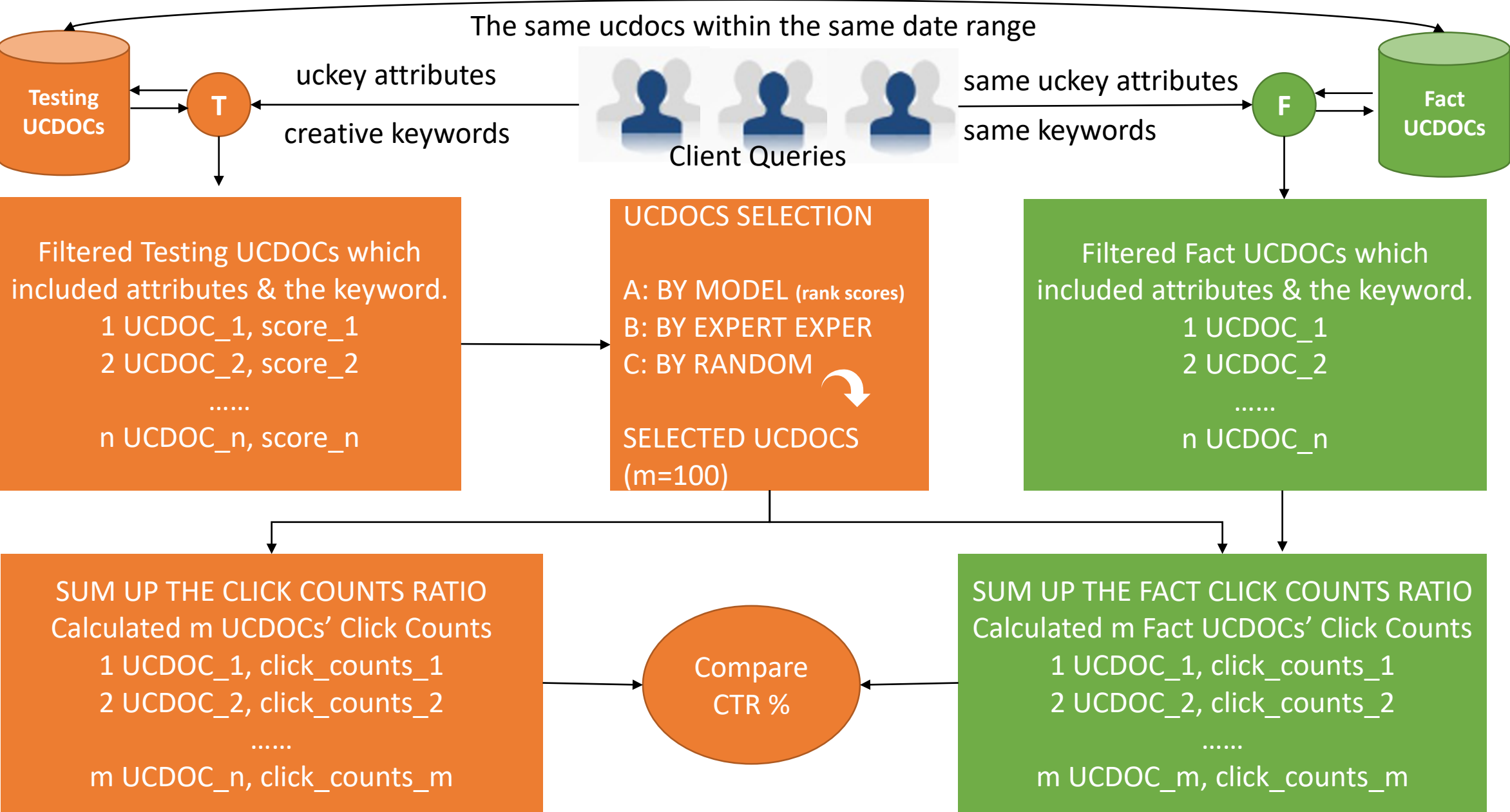
e.g. 27,559 GUCDOCs.

[Row(region_id=26, age=1, gender=0, net_type=u'4G', media_category=u'Huawei Reading', media=u'native', uckey=u'native,Huawei Reading,4G,0,1,26', interval_starting_time=[1579392000], keyword_indexes=[u'26'], keyword_indexes_click_counts=[u'26:0'], keyword_indexes_show_counts=[u'26:1'], keywords=[u'social'], keywords_click_counts=[u'social:0'], keywords_show_counts=[u'social:1'], uckey_index=7393, media_index=1, media_category_index=5, net_type_index=3, gender_index=1, age_index=1, region_id_index=26)

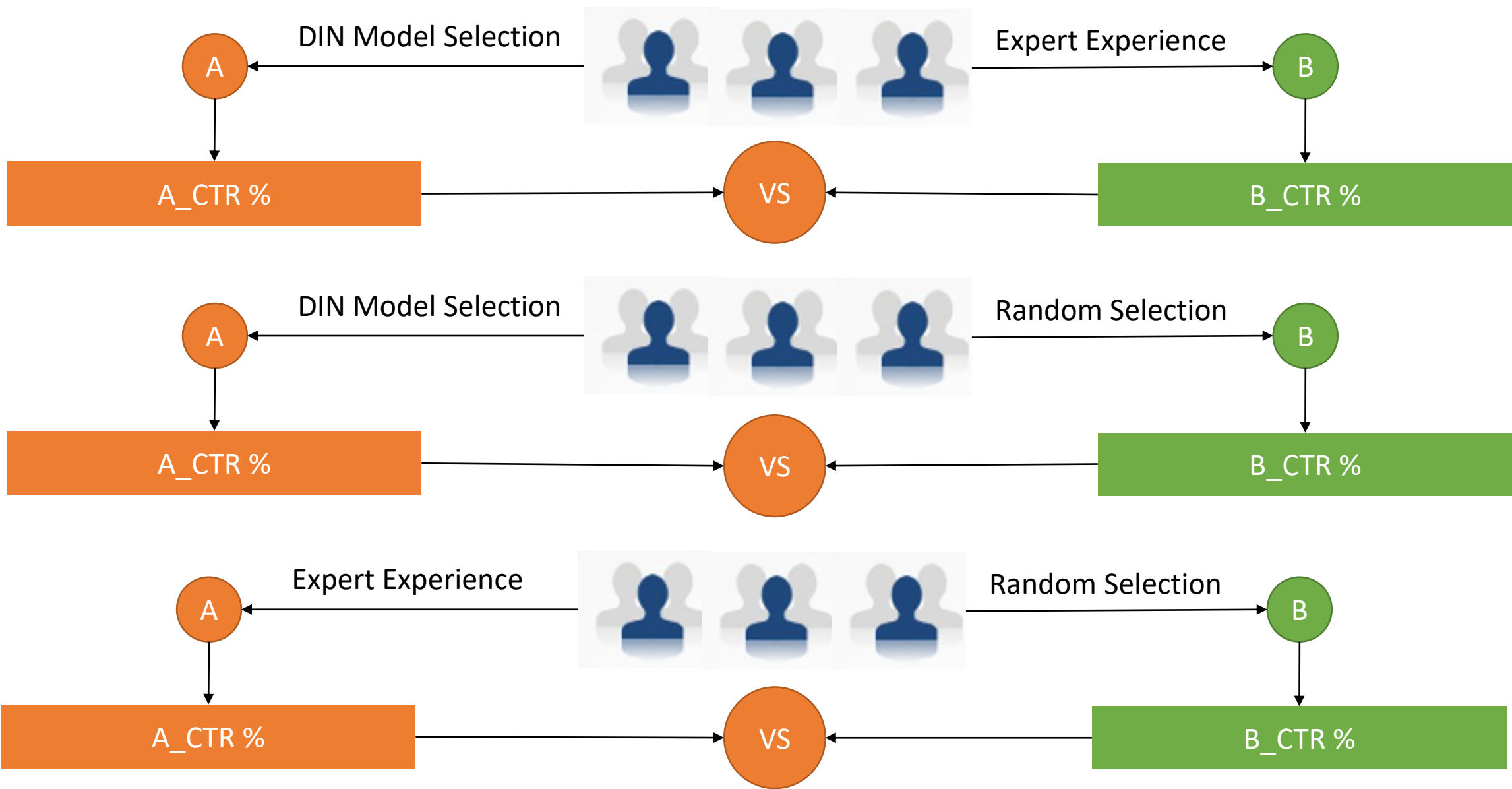
e.g. 7,885,836 logs.

[Row(is_click=0, action_time=u'2020-03-12 21:17:59', keyword=u'shopping', keyword_index=25, media=u'native', media_category=u'Huawei Brower', net_type=u'3G', gender=1, age=3, adv_id=u'45018971', action_time_seconds=1584073079, interval_starting_time=1584057600, uckey=u'native,Huawei Brower,3G,1,3,69', region_id=69),

Testing Processing Logic Design



Testing Processing Logic Design: A/B Testing



Testcases Design Rules

- Based on the specific attributes and creative keywords.
- Three methods of selecting ucdocs: by din model, by expert experience, by random selection.

Testing Result with testcase designed

Dates: starting_date: '2019-12-19', ending_date: '2020-04-15', cutting_date: '2020-03-20'
History days: 90, Testing days: 28, History keywords length: 5.
Expert experience: gender: female, age: 3 for testing keyword 'living-makeup' and 'shopping'.
Total ucdocs number: 25798, in history days: 25422, in testing days: 21508 (84% of history).
Ucdocs number selected by din model and random selection: 2500 (10%).



Forecasting CTR by DIN model:			
keyword	keyword_click_count	keyword_show_count	CTR
shopping	2392	78288	0.030553853

Forecasting CTR by random selection:			
random	keyword_click_count	keyword_show_count	CTR
	345	41415	0.008330315

Forecasting CTR by expert experience:			
'gender': female, 'age': 30	keyword_click_count	keyword_show_count	CTR
	261	23561	0.0110776285

Forecasting CTR by DIN model:			
keyword	keyword_click_count	keyword_show_count	CTR
living-makeup	22	958	0.02296451

Forecasting CTR by random selection:			
random	keyword_click_count	keyword_show_count	CTR
	2	701	0.0028530671

Forecasting CTR by expert experience:			
'gender': female, 'age': 30	keyword_click_count	keyword_show_count	CTR
	14	1535	0.009120521

5000

By male:

Forecasting CTR by expert experience:			
keyword	keyword_click_count	keyword_show_count	CTR
living-makeup	0	20	0.0

By female:

Forecasting CTR by expert experience:			
keyword	keyword_click_count	keyword_show_count	CTR
living-makeup	14	1535	0.009120521

Testing Result: Keywords' real CTRs in total and testing datasets.

Keyword's CTR distribution in 118 days (total)

keyword	keyword_click_count	keyword_show_count	CTR
social	16757	640039	0.026181217
video	67582	3338385	0.02024392
info	33734	1696266	0.01988721
health	6	362	0.016574586
travel	324	22778	0.014224252
game-fishing	27	2033	0.013280866
living-house	278	23030	0.012071212
living-map	59	4941	0.011940903
shopping	22682	2040557	0.011115593
other	56	5169	0.010833817
game-rpg	92	8628	0.010662958
game-act	11	1057	0.010406812
education	211	23370	0.009028669
game-cnc	8	905	0.008839779
living-makeup	153	17659	0.008664138
living-mon	81	9874	0.008203362
game-slg	45	6351	0.0070854984
reading	168	24652	0.006814863
entertainment	94	19678	0.004776908
game-avg	0	5	0.0
game-moba	0	1	0.0
sports	0	96	0.0

Keyword's CTR distribution in 28 days (testing)

keyword	keyword_click_count	keyword_show_count	CTR
social	2685	106372	0.025241604
video	17432	866468	0.020118458
info	3013	170497	0.017671864
game-rpg	10	669	0.014947684
game-fishing	23	1868	0.012312634
living-mon	24	1994	0.012036108
shopping	4268	445175	0.009587241
living-house	27	3447	0.007832898
education	89	11952	0.0074464525
living-makeup	29	3910	0.0074168798
reading	77	13073	0.0058900025
game-cnc	1	195	0.0051282053
entertainment	0	1	0.0
game-slg	0	23	0.0
other	0	85	0.0
sports	0	20	0.0
travel	0	3	0.0

The testing data has similar ranks and ratios of CTR on top keywords. The two keywords of 'shopping and 'living-makeup' have very close CTR results in real data and the expert experience forecasting.

Testing Result: UCDOCS' CTR Analysis

- Testing fact dataset: '2020-03-20' - '2020-04-15', 28 days.
- 21508 UCDOCS, 17 keywords, 50650 UCDOC-keyword pairs. (97% pairs have 0 click counts.)
- Average CTR is 1%, Avg CTR of non-zero is 37%.
- CTR list with highest click counts.

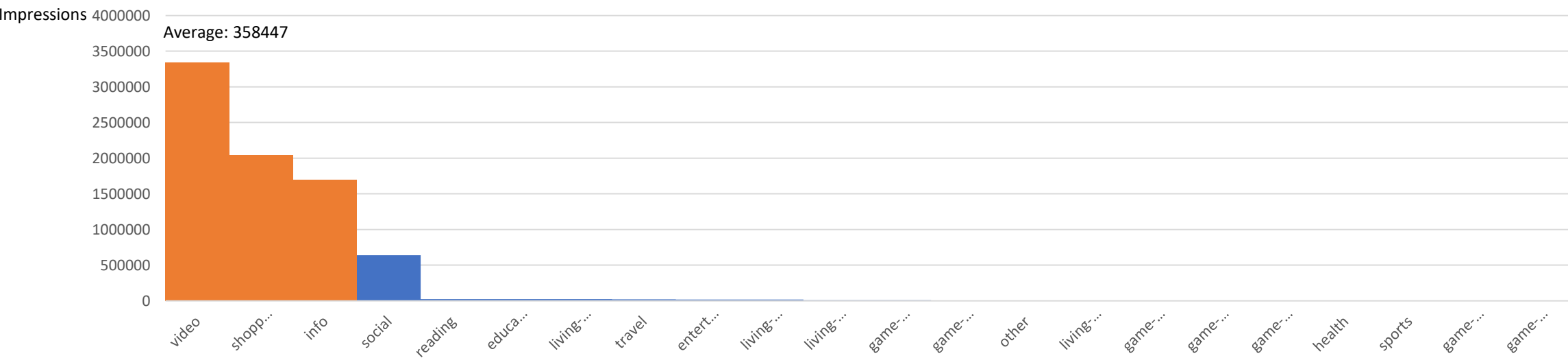
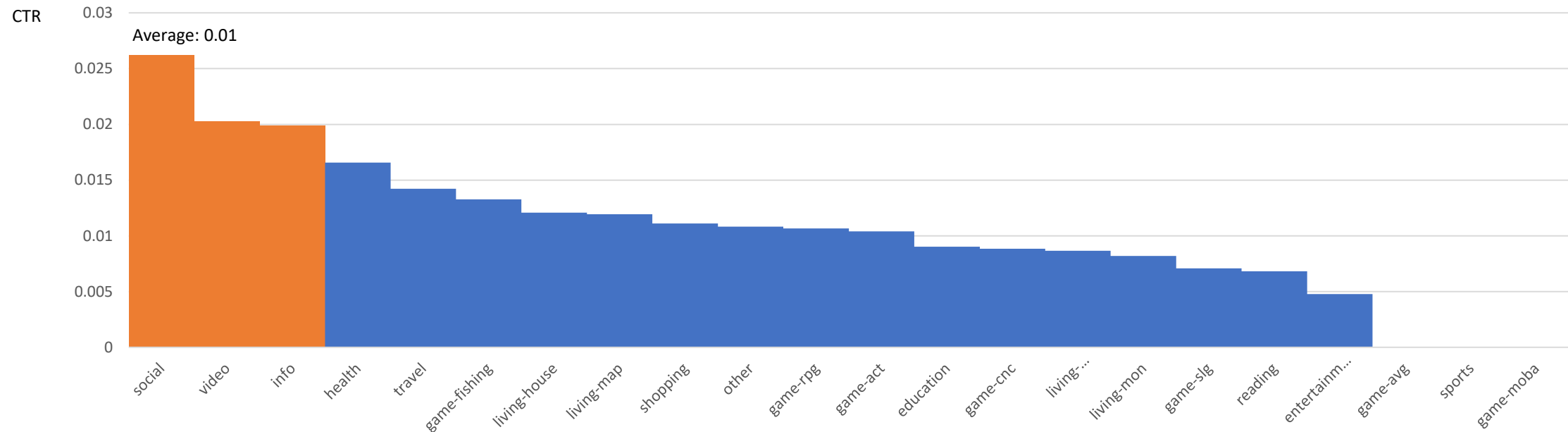
uckey	keyword_index	keyword	keyword_click_count	keyword_show_count	CTR
roll,Huawei Video,WIFI,0,6,82	29	video	1447	2624	0.55144817
roll,Huawei Video,WIFI,0,5,82	29	video	1429	2566	0.5568979
roll,Huawei Video,WIFI,0,4,82	29	video	985	1784	0.55213004
roll,Huawei Video,4G,0,5,82	29	video	772	1045	0.738756
roll,Huawei Video,WIFI,0,2,82	29	video	479	892	0.53699553
roll,Huawei Video,WIFI,0,3,82	29	video	466	766	0.6083551
native,Huawei Brower,WIFI,0,5,82	29	video	442	3167	0.13956426
roll,Huawei Video,WIFI,1,4,82	29	video	429	570	0.7526316
native,Huawei Brower,WIFI,0,4,82	29	video	380	3777	0.10060895
splash,Huawei Reading,WIFI,1,2,82	29	video	356	1032	0.34496123

- CTR list with highest show counts.

uckey	keyword_index	keyword	keyword_click_count	keyword_show_count	CTR
native,Huawei Brower,WIFI,0,4,82	29	video	380	3777	0.10060895
native,Huawei Brower,WIFI,0,5,82	29	video	442	3167	0.13956426
native,Huawei Video,WIFI,0,5,82	29	video	285	2980	0.09563758
native,Huawei Brower,WIFI,0,4,80	29	video	0	2972	0.0
native,Huawei Brower,WIFI,0,3,82	29	video	248	2907	0.085311316
native,Huawei Video,WIFI,0,5,80	29	video	0	2828	0.0
native,Huawei Brower,4G,0,4,82	25	shopping	166	2667	0.06224222
roll,Huawei Video,WIFI,0,6,82	29	video	1447	2624	0.55144817
native,Huawei Brower,WIFI,0,5,80	29	video	0	2579	0.0
roll,Huawei Video,WIFI,0,5,82	29	video	1429	2566	0.5568979

Testing Result: Keywords' CTRs & impressions analysis

Top keywords take higher CTRs. A few bottom keywords take 0 CTRs. **The average CTR is 1%.**
A few top keywords take most of the impressions. **18% keywords take 98% impressions.**



Thank you!