#	△pub	Team Name	Team Members	Score	Entries	Last
1	—	¯_(ツ)_/¯		0.47550	6	2y
2	—	(ಠ_ಠ)		0.47968	51	2y
3	—	**Novice**		0.48330	36	2y
4	—	:-)		0.49123	26	2y
5	—	**KW Wu**		0.49504	83	2y
6	—	**Eureka**		0.49701	34	2y
7	—	**Dmitry Ulyanov**		0.50011	37	2y

1. Problem statement
2. Data format
3. Models
4. Feature engineering

Data

In [20]:
```
train[:8]
```

Out[20]:

	TripType	VisitNumber	Weekday	Upc	ScanCount	DepartmentDescription	FinelineNumber
0	999	5	Friday	68113152929	-1	FINANCIAL SERVICES	1000
1	30	7	Friday	60538815980	1	SHOES	8931
2	30	7	Friday	7410811099	1	PERSONAL CARE	4504
3	26	8	Friday	2238403510	2	PAINT AND ACCESSORIES	3565
4	26	8	Friday	2006613744	2	PAINT AND ACCESSORIES	1017
5	26	8	Friday	2006618783	2	PAINT AND ACCESSORIES	1017
6	26	8	Friday	2006613743	1	PAINT AND ACCESSORIES	1017
7	26	8	Friday	7004802737	1	PAINT AND ACCESSORIES	2802

In [24]:
```
len(train), train.VisitNumber.nunique()
```

Out[24]:
```
(647054, 95674)
```

In [25]:
```
for c in train:
    print c, train[c].nunique()
```

```
TripType 38
VisitNumber 95674
Weekday 7
Upc 97714
ScanCount 39
DepartmentDescription 68
FinelineNumber 5195
```

Default data format

```
In [15]: train[train.VisitNumber == 8]
```

Out[15]:

	TripType	VisitNumber	Weekday	Upc	ScanCount	DepartmentDescription	FinelineNumber
3	26	8	Friday	2238403510	2	PAINT AND ACCESSORIES	3565
4	26	8	Friday	2006613744	2	PAINT AND ACCESSORIES	1017
5	26	8	Friday	2006618783	2	PAINT AND ACCESSORIES	1017
6	26	8	Friday	2006613743	1	PAINT AND ACCESSORIES	1017
7	26	8	Friday	7004802737	1	PAINT AND ACCESSORIES	2802
8	26	8	Friday	2238495318	1	PAINT AND ACCESSORIES	4501
9	26	8	Friday	2238400200	-1	PAINT AND ACCESSORIES	3565
10	26	8	Friday	5200010239	1	DSD GROCERY	4606
11	26	8	Friday	88679300501	2	PAINT AND ACCESSORIES	3504
12	26	8	Friday	22006000000	1	MEAT - FRESH & FROZEN	6009
13	26	8	Friday	2236760452	1	PAINT AND ACCESSORIES	7

```
In [14]: ' '.join(train[train.VisitNumber == 8].DepartmentDescription.astype(str))
```

```
Out[14]: 'PAINT AND ACCESSORIES PAINT AND ACCESSORIES PAINT AND ACCESSORIES PAINT AND ACCESSORIES PAINT AND ACCESSORIES PAINT AND ACCESS
ORIES PAINT AND ACCESSORIES DSD GROCERY PAINT AND ACCESSORIES MEAT - FRESH & FROZEN PAINT AND ACCESSORIES PAINT AND ACCESSORIES
PAINT AND ACCESSORIES PAINT AND ACCESSORIES PAINT AND ACCESSORIES DAIRY PETS AND SUPPLIES PAINT AND ACCESSORIES HOUSEHOLD CHEMI
CALS/SUPP PAINT AND ACCESSORIES PAINT AND ACCESSORIES PETS AND SUPPLIES nan'
```

Processed data format

```
gb = data.groupby('VisitNumber')['DepartmentDescription'].value_counts()
gb[:9]
```

```
VisitNumber  DepartmentDescription
5            FINANCIAL SERVICES            1
7            PERSONAL CARE                 1
             SHOES                         1
8            PAINT AND ACCESSORIES        16
             PETS AND SUPPLIES             2
             DAIRY                         1
             DSD GROCERY                   1
             HOUSEHOLD CHEMICALS/SUPP      1
             MEAT - FRESH & FROZEN         1
Name: DepartmentDescription, dtype: int64
```

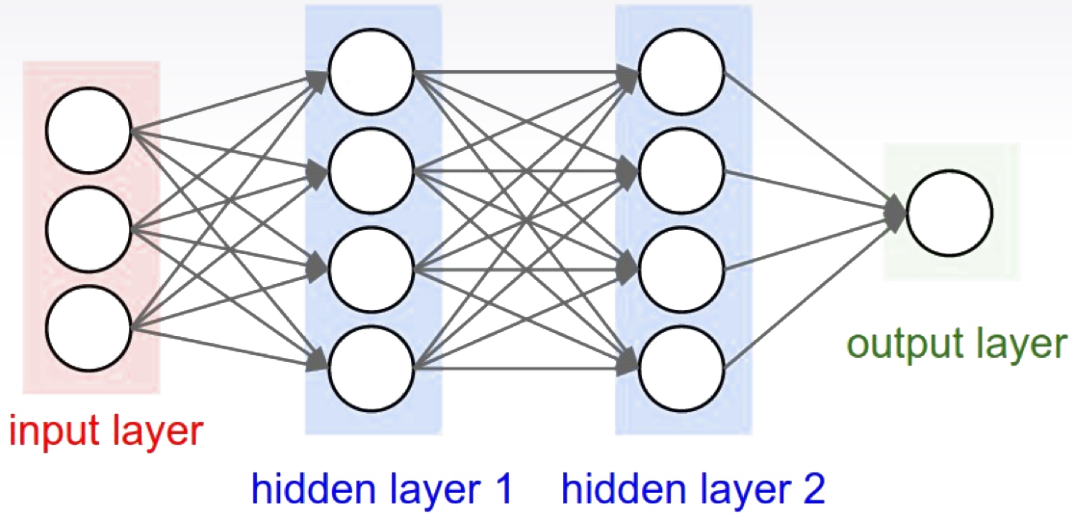
```
gb = gb.unstack()
gb[:3]
```

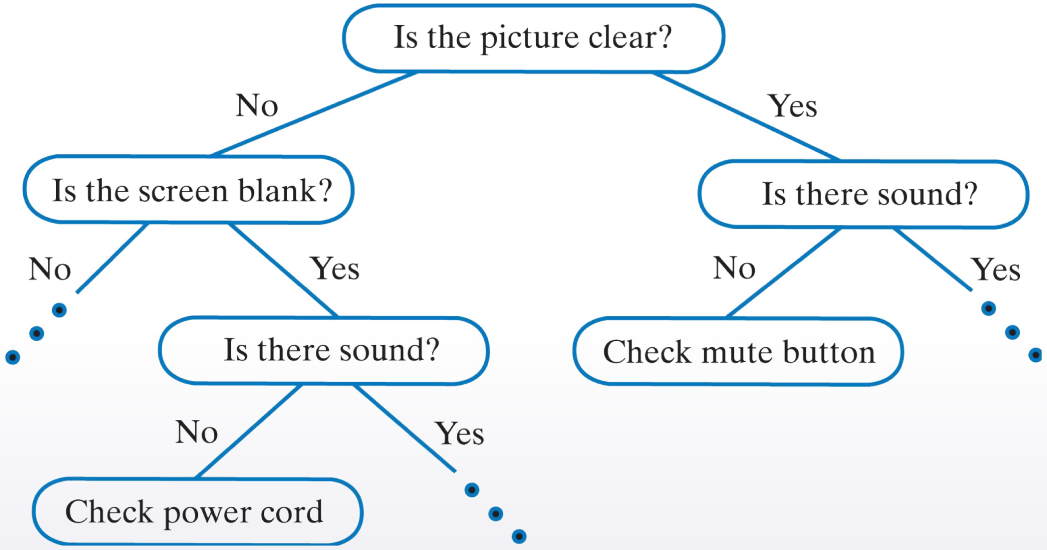
DepartmentDescription	ACCESSORIES	BAKERY	BOYS WEAR	CANDY, TOBACCO, COOKIES	COMM BREAD	COOK AND DINE	DAIRY	DSD GROCERY	FABRICS AND CRAFTS	FINANCIAL SERVICES	...
VisitNumber											
5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	...
7	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
8	NaN	NaN	NaN	NaN	NaN	NaN	1.0	1.0	NaN	NaN	...

3 rows × 28 columns

Models

Models: interactions



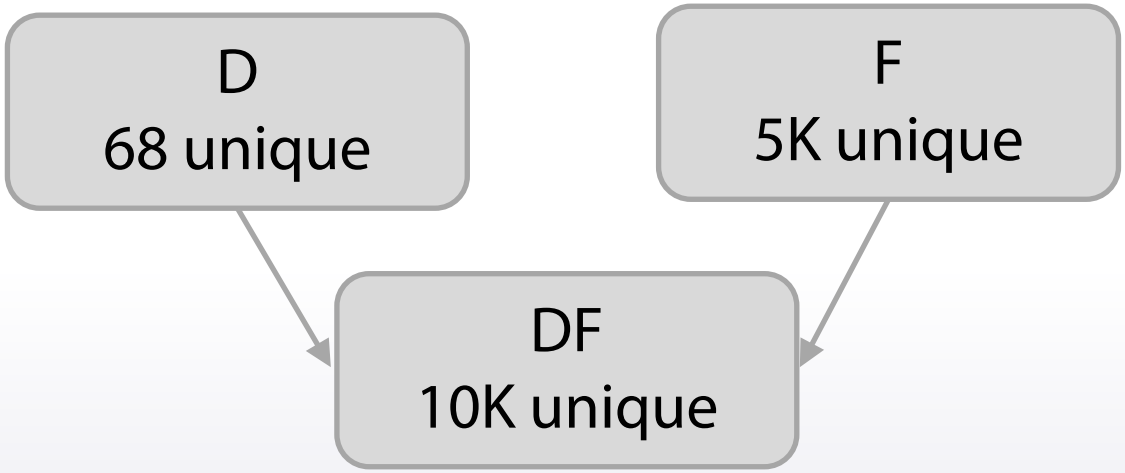


Feature engineering

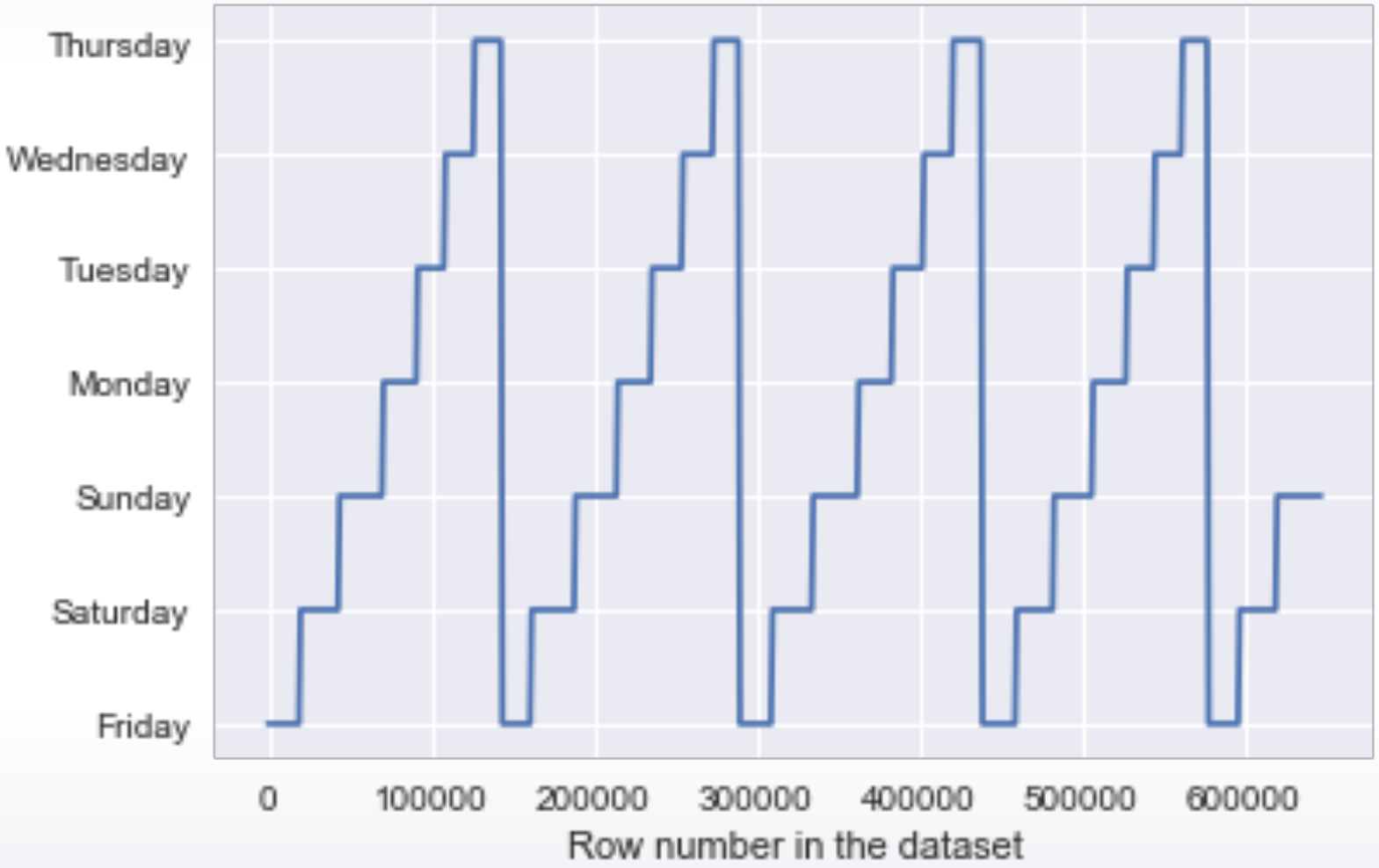
```
In [20]: train[:8]
```

Out[20]:

	TripType	VisitNumber	Weekday	Upc	ScanCount	DepartmentDescription	FinelineNumber
0	999	5	Friday	68113152929	-1	FINANCIAL SERVICES	1000
1	30	7	Friday	60538815980	1	SHOES	8931
2	30	7	Friday	7410811099	1	PERSONAL CARE	4504
3	26	8	Friday	2238403510	2	PAINT AND ACCESSORIES	3565
4	26	8	Friday	2006613744	2	PAINT AND ACCESSORIES	1017
5	26	8	Friday	2006618783	2	PAINT AND ACCESSORIES	1017
6	26	8	Friday	2006613743	1	PAINT AND ACCESSORIES	1017
7	26	8	Friday	7004802737	1	PAINT AND ACCESSORIES	2802



Feature engineering




Walmart


Walmart

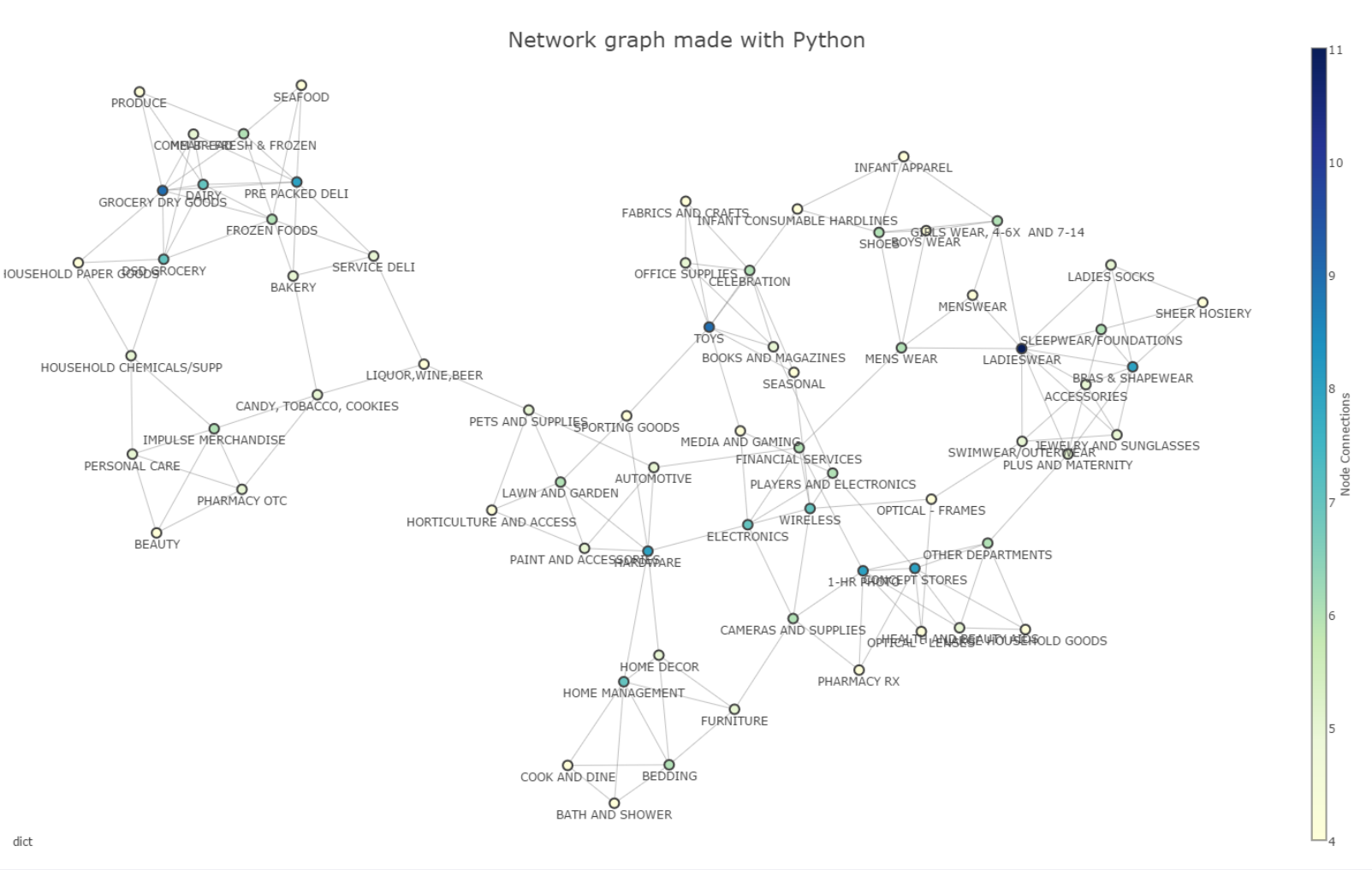
EXTRA

Besides...

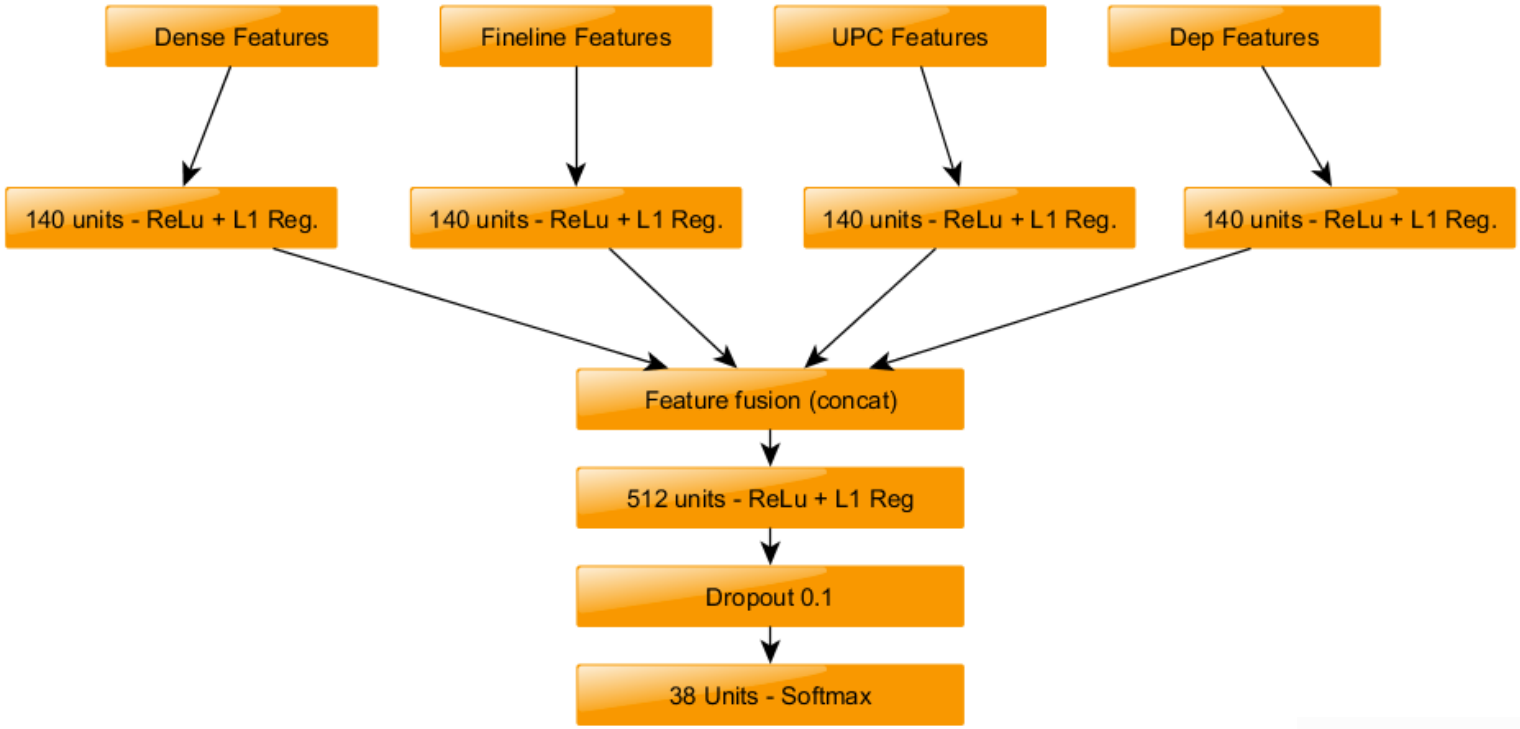
```
In [18]: train.DepartmentDescription.unique()

Out[18]: array(['FINANCIAL SERVICES', 'SHOES', 'PERSONAL CARE',
                'PAINT AND ACCESSORIES', 'DSD GROCERY', 'MEAT - FRESH & FROZEN',
                'DAIRY', 'PETS AND SUPPLIES', 'HOUSEHOLD CHEMICALS/SUPP', nan,
                'IMPULSE MERCHANDISE', 'PRODUCE', 'CANDY, TOBACCO, COOKIES',
                'GROCERY DRY GOODS', 'BOYS WEAR', 'FABRICS AND CRAFTS',
                'JEWELRY AND SUNGLASSES', 'MENS WEAR', 'ACCESSORIES',
                'HOME MANAGEMENT', 'FROZEN FOODS', 'SERVICE DELI',
                'INFANT CONSUMABLE HARDLINES', 'PRE PACKED DELI', 'COOK AND DINE',
                'PHARMACY OTC', 'LADIESWEAR', 'COMM BREAD', 'BAKERY',
                'HOUSEHOLD PAPER GOODS', 'CELEBRATION', 'HARDWARE', 'BEAUTY',
                'AUTOMOTIVE', 'BOOKS AND MAGAZINES', 'SEAFOOD', 'OFFICE SUPPLIES',
                'LAWN AND GARDEN', 'SHEER HOSIERY', 'WIRELESS', 'BEDDING',
                'BATH AND SHOWER', 'HORTICULTURE AND ACCESS', 'HOME DECOR', 'TOYS',
                'INFANT APPAREL', 'LADIES SOCKS', 'PLUS AND MATERNITY',
                'ELECTRONICS', 'GIRLS WEAR, 4-6X  AND 7-14', 'BRAS & SHAPEWEAR',
                'LIQUOR,WINE,BEER', 'SLEEPWEAR/FOUNDATIONS', 'CAMERAS AND SUPPLIES',
                'SPORTING GOODS', 'PLAYERS AND ELECTRONICS', 'PHARMACY RX',
                'MENSWEAR', 'OPTICAL - FRAMES', 'SWIMWEAR/OUTERWEAR',
                'OTHER DEPARTMENTS', 'MEDIA AND GAMING', 'FURNITURE',
                'OPTICAL - LENSES', 'SEASONAL', 'LARGE HOUSEHOLD GOODS',
                '1-HR PHOTO', 'CONCEPT STORES', 'HEALTH AND BEAUTY AIDS'], dtype=object)
```

DepartmentDescription connection graph



Interactions! Neural nets



@Tarantula

Links

Read the forum for good code & solutions write-ups!