# Dell Customer Service Data Analysis

ZHENWEN PAN

# Overview

- Here is the general layout of this presentation
- Analysis on Products and Customer Services:
  - Initial Data QC and Data Cleansing
  - Data Exploration and Analysis (focusing on products and services)
  - Observation, Conclusion and More Questions
- Exploring Graph DB with the Same Data Set

# Overview

- Here is the general layout of this presentation

- **Analysis on Products and Customer Services:**
  - **Initial Data QC and Data Cleansing**
  - Data Exploration and Analysis (focusing on products and services)
  - Observation, Conclusion and More Questions

- Exploring Graph DB with the Same Data Set

# First Glance at the Data Table

- For the raw input table, 99999 instances and 19 columns are included (data size: 15mb)
- For the data type, both categorical and numerical type data are included:
  - For the **categorical** data, the main information is mainly about **product types, instance/product region, product issues/topics, service and fixes (diagnostics and parts)**
  - For the **numerical** data, except for the **assistant ID, manufacturing and warranty time(year + week), sent part count and agent tenure days**
  - The service data all happened in the same week (**the 40th week of 2018**)

# First Glance at the Data Table (Continued)

- Main issues for the data:
  - Duplications
  - Missing values (missing values are the largest proportion in some of the columns)
  - For the manufacturing and warranty time data, the information is not accurate (i.e., warranty started earlier than the manufacturing time)
- Each individual issue is to be QC and addressed

# Other Initial QC

- Data type:

```
df.dtypes
```

```
asst_id                int64
product_type          object
mnfture_wk             int64
contract_st            int64
contract_end           int64
contact_wk             int64
contact_type          object
issue_type            object
topic_category        object
parts_sent            object
repair_type           object
repeat_ct              int64
parts_ct               int64
agent_tenure_indays  float64
contact_manager_flg    int64
diagnostics           object
repeat_parts_sent     object
region                object
country               object
repetition             int64
contract_dur_wk        int64
mnf_contract_wk        int64
contact_contract_wk    int64
mn_year                int64
contract_st_year       int64
contact_year           int64
dtype: object
```

- Summary for Categorical Data

```
Number of categories for " product_type " : 3
Laptops              57444
Desktops             18719
Other Electronics     2434
NaN                     15
Name: product_type, dtype: int64

Number of categories for " contact_type " : 4
VOICE      66170
CHAT       11304
EMAIL       1044
UNKNOWN       94
Name: contact_type, dtype: int64

Number of categories for " issue_type " : 80
NaN                              11358
System Board Components          10188
Fee Based Support                 8177
Operating System                  7550
Hard Drive                        6759
                                 ...
Shipping Error / Factory Error       1
Compellent - Hardware                1
Lost / Stolen in Transit             1
Order Tracking                       1
MWD                                  1
Name: issue_type, Length: 81, dtype: int64

Number of categories for " topic_category " : 63
NaN                    25983
Booting                12095
LCD/Monitor             4194
Call Logs               3759
General Queries         3330
```

- Summary for Numerical Data

```
inp_raw.describe()
```

| | asst_id | mnfture_wk | contract_st | contract_end | contact_wk | repeat_ct | parts_ct | agent_tenure_indays | contact_manager_flg |
|---|---|---|---|---|---|---|---|---|---|
| count | 100000.000000 | 100000.000000 | 100000.000000 | 100000.000000 | 100000.0 | 100000.000000 | 100000.00000 | 99986.000000 | 100000.000000 |
| mean | 35702.117890 | 201672.108910 | 201671.595910 | 201951.897740 | 201840.0 | 0.074090 | 1.15955 | 985.886844 | 0.012430 |
| std | 25038.092288 | 156.139692 | 158.940598 | 186.777325 | 0.0 | 0.262491 | 1.56942 | 862.044503 | 0.110795 |
| min | 0.000000 | 201001.000000 | 200547.000000 | 201006.000000 | 201840.0 | 0.000000 | 0.00000 | 218.000000 | 0.000000 |
| 25% | 12382.750000 | 201612.000000 | 201612.000000 | 201903.000000 | 201840.0 | 0.000000 | 0.00000 | 536.000000 | 0.000000 |
| 50% | 32898.500000 | 201719.000000 | 201720.000000 | 201938.000000 | 201840.0 | 0.000000 | 1.00000 | 801.000000 | 0.000000 |
| 75% | 57701.250000 | 201808.000000 | 201809.000000 | 202045.000000 | 201840.0 | 0.000000 | 2.00000 | 1018.000000 | 0.000000 |
| max | 82441.000000 | 201927.000000 | 202630.000000 | 202953.000000 | 201840.0 | 3.000000 | 51.00000 | 8922.000000 | 1.000000 |

# Duplications

- Duplications were handled in two stages:
  - Simply removing the obvious duplicated instances
  - Further removing duplications with same product and services information but only different country/region information (or one record with country and the other one without)
  - There may be ways to further reduce the instance number, but the safest way is to only isolate the country/region columns

# Duplications

- Step QC and actions after removing duplications:
  - After first stage removal. Instance number reduced to 86536

```
86536 rows × 19 columns
```

  - In the second stage, duplications were removed one more time after dropping off the country and region columns. After this operation, the instance number further were reduced to 83172

```
83172 rows × 19 columns
```

  - The country/region columns were merged back afterwards
  - An instance repetition attribute/column was created to keep track of the repeated records

**This summary is after dropping some records with time issue and so total number is smaller than the one above. Will mention this in the later slide**

```
collections.Counter(df_edit['repetition'])
```

```
Counter({1: 77361, 2: 1190, 6: 6, 7: 7, 3: 36, 4: 12})
```

# Missing Values

- There were quite some missing values in the table. Especially for the following information:
  - Issue Type
  - Topics

```
df_edit['issue_type'].count(), df_edit['topic_category'].count()
(67254, 52629)
```

**The total record number is ~81000**

  - Sent part and resent part (But it makes sense since lot of fixes didn't need to sent parts)
  - Imputation doesn't seem to be easy for these columns and 'unknown' was added to replace the missing values for these columns

# Manufacturing and Warranty Weeks

- The manufacturing week and warranty (start/end) weeks were explored.
- Some situations of in inaccuracy were observed:
  - Warranty started later than customer contact week (729 records)

```
negative value count: 729
-406 -297 -39 -38 -37 -36 -35 -34 -33 -32 -31 -30 -29 -28 -27 -26 -25 -24 -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 -11 -
10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 7
8 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 11
```

  - Manufacturing week was later than customer contact week (1243 records)

```
negative value count: 1243
-39 -38 -37 -36 -35 -34 -33 -32 -31 -30 -29 -28 -27 -26 -25 -24 -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 -11 -10 -9 -8 -
7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 8
1 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117
```

  - Warranty started earlier than manufacturing week (3861 records)

```
negative value count: 3861
-524 -493 -484 -458 -456 -447 -431 -372 -371 -369 -361 -358 -352 -334 -327 -313 -291 -288 -272 -269 -265 -262 -257 -253 -252 -2
50 -249 -248 -243 -240 -233 -232 -230 -227 -226 -225 -222 -221 -219 -215 -214 -213 -212 -210 -208 -207 -206 -205 -204 -203 -202
-201 -199 -197 -195 -194 -193 -191 -190 -189 -186 -184 -183 -182 -181 -180 -179 -177 -173 -172 -170 -169 -168 -167 -166 -165 -1
63 -162 -161 -160 -159 -158 -157 -156 -155 -154 -153 -152 -151 -150 -149 -148 -147 -146 -145 -144 -143 -142 -141 -140 -139 -138
-137 -136 -135 -134 -133 -132 -131 -130 -129 -128 -127 -126 -125 -124 -123 -122 -121 -120 -119 -118 -117 -116 -115 -114 -113 -1
12 -111 -109 -108 -107 -106 -105 -104 -103 -102 -101 -100 -98 -97 -96 -95 -94 -93 -92 -91 -90 -89 -88 -87 -86 -85 -84 -83 -82 -
81 -80 -79 -78 -77 -76 -75 -74 -73 -72 -71 -70 -69 -68 -67 -66 -65 -64 -63 -62 -61 -60 -59 -58 -57 -56 -55 -54 -53 -52 -51 -50
-49 -48 -47 -46 -45 -44 -43 -42 -41 -40 -39 -38 -37 -36 -35 -34 -33 -32 -31 -30 -29 -28 -27 -26 -25 -24 -23 -22 -21 -20 -19 -18
-17 -16 -15 -14 -13 -12 -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
```

# Manufacturing and Warranty Weeks

- The manufacturing week and warranty (start/end) weeks were explored.
- Some situations of in inaccuracy were observed:
  - Warranty started later than customer contact week (729 records)
  - Manufacturing week was later than customer contact week (1243 records)
  - Warranty started earlier than manufacturing week (3861 records)

- Some of the discrepancy are not even small (i.e., warranty started 406 weeks after customer contacting for issues)
- The correct information is hard to inferred. Especially when the discrepancy is large
- Since number of these problematic records is less than 5%, we can drop them especially when doing some time related analysis
- The total number of instance is **78612** after dropping those problematic records

# Header Created for Convenience

- Instance Repetition (mentioned in the previous slides)
- Manufacturing, contract start year, contract end year and customer contact year(all 2018 though)
- Weeks of issue happening after warranty start (**how soon the problem happened** after the products being used)
- Agent tenure in days group (The days were grouped by every 2000 days)

# Overview

- Here is the general layout of this presentation

- **Analysis on Products and Customer Services:**
  - Initial Data QC and Data Cleansing
  - **Data Exploration and Analysis (focusing on products and services)**
  - Observation, Conclusion and More Questions

- Exploring Graph DB with the Same Data Set

# Data Exploration and Analysis

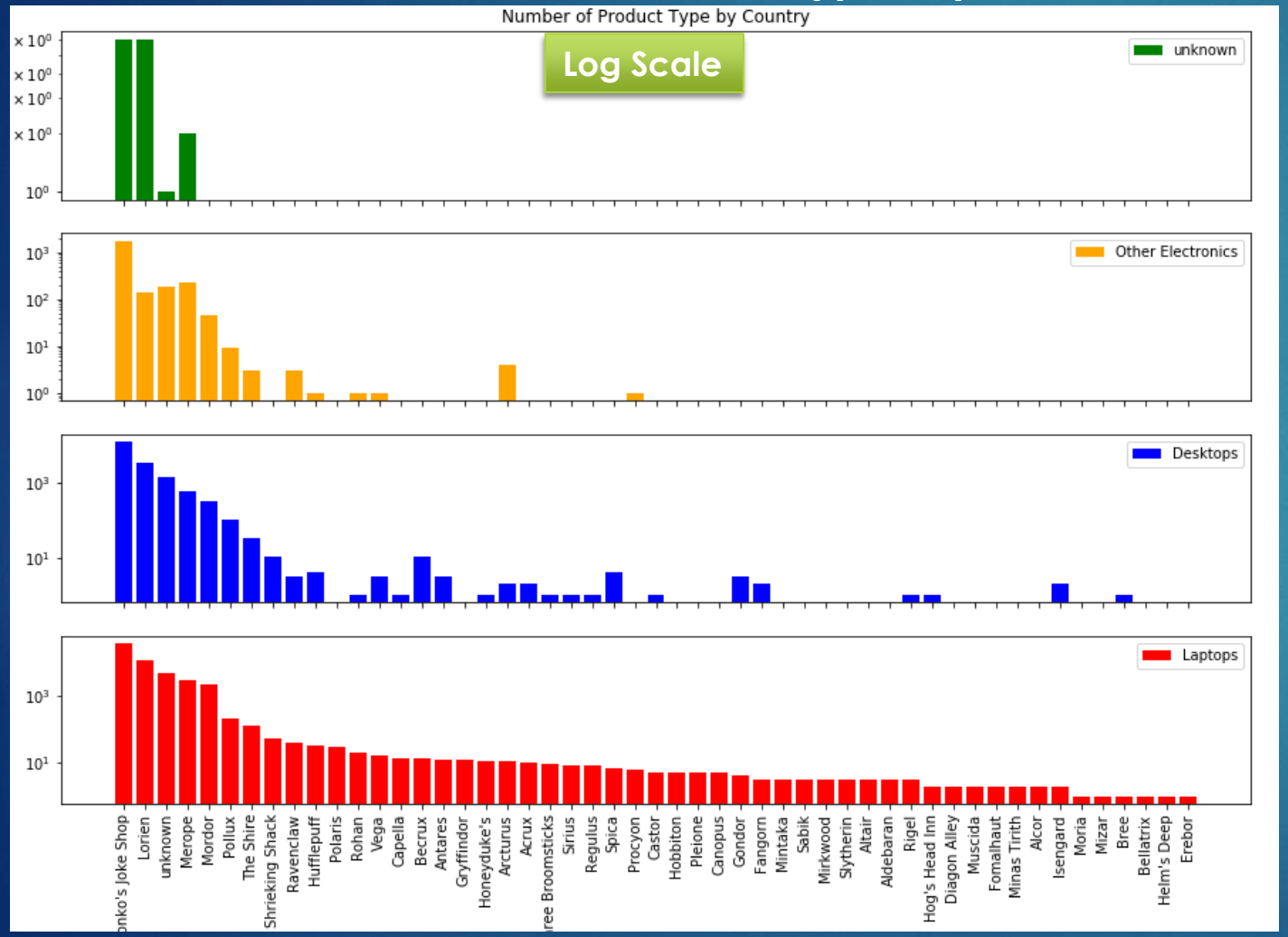
- The are two focuses for this analysis: **Product** and **Services**
  - For **product analysis**, here are aspect for the analysis:
    - Instances counts/distributions by product type, country and region
    - Extract top issues (identified by customers) and topics (identified by tech support agents) segmented by product type, region
    - Product type issue county by year and how soon the products having issue after warranty start
  - For the **service/fixing analysis**, several aspects were also explored:
    - Top issues and topics reported by different contact type
    - Top issues and topics handled by different repair types (Hard of Soft)
    - Top issues involving managers
    - Parts:
      - Top sent/resent parts to be replaced for the issues/fixing
      - Year distribution for the top sent/resent parts and **how soon** needing replacement after warranty started
    - Something about the agents: Diagnostic usages, number of sent parts and contacting manager flag vs. agent tenure in days

# Product

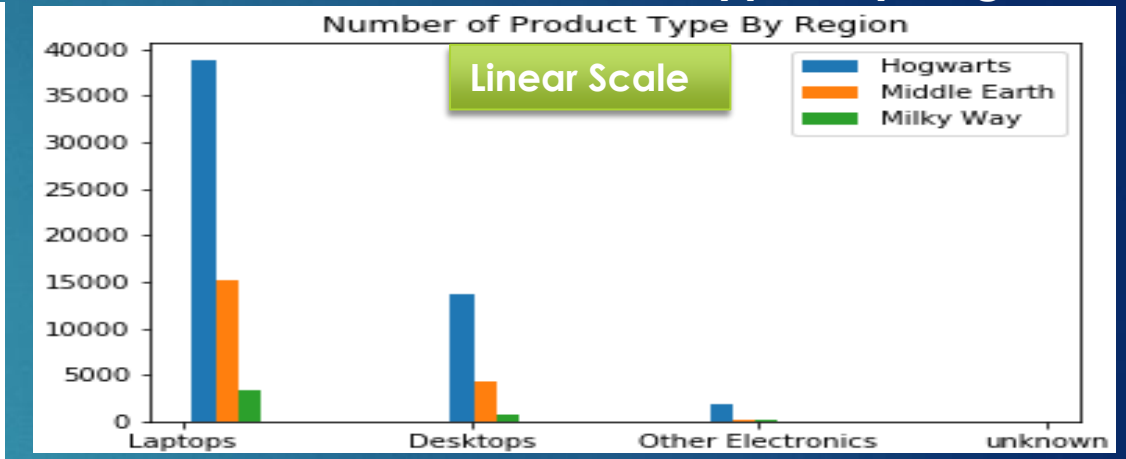
- For **product analysis**, here are aspect for the analysis:
  - Instances counts/distributions by product type, country and region
  - Extract top issues (identified by customers) and topics (identified by tech support agents) segmented by product type, region
  - Product type issue county by year and how soon the products having issue after warranty start
- Some notes for the charts for the analysis:
  - Word clouds are introduced to visualize the most prominent issues/topics for different segmentation at the first place
  - Some charts (mostly for region and country comparisons) will include both linear and log scales since some of the segments are way large than the others
  - The bar chart for top issue/topics/parts will have both count and in-group proportion on the same chart

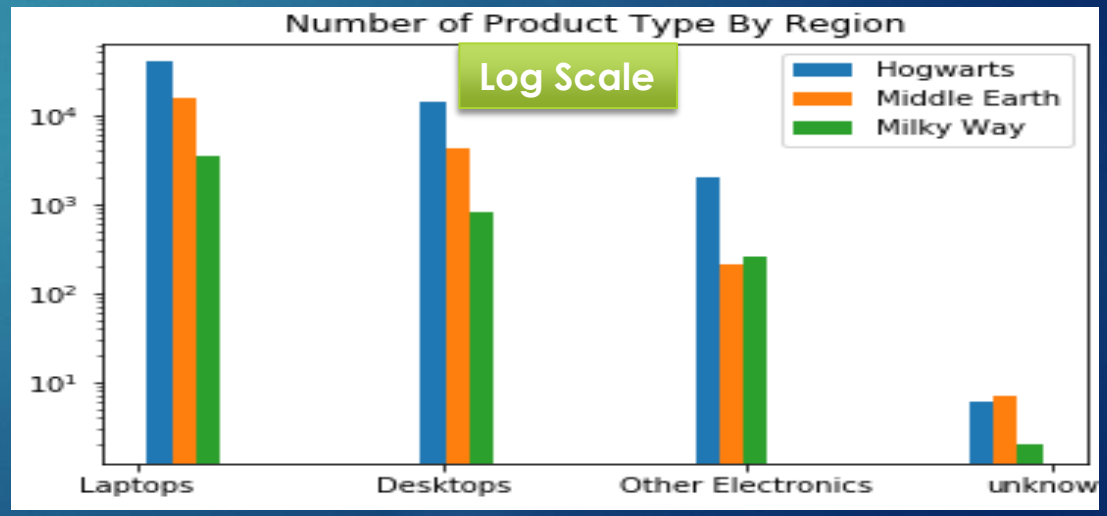
# Number of Product Type by Region

## Counts for Different Product Types by Countries




## Counts for Different Product Types by Regions




**Laptop is the largest products/issues among all the sale countries and regions (about 2.5 times of the Desktops )

# Issue and Topic by Product Type

- Product types:
  - Laptop
  - Desktop
  - Other Electronics
- Items included:
  - Word Cloud
  - Bar Charts including both counts and proportions

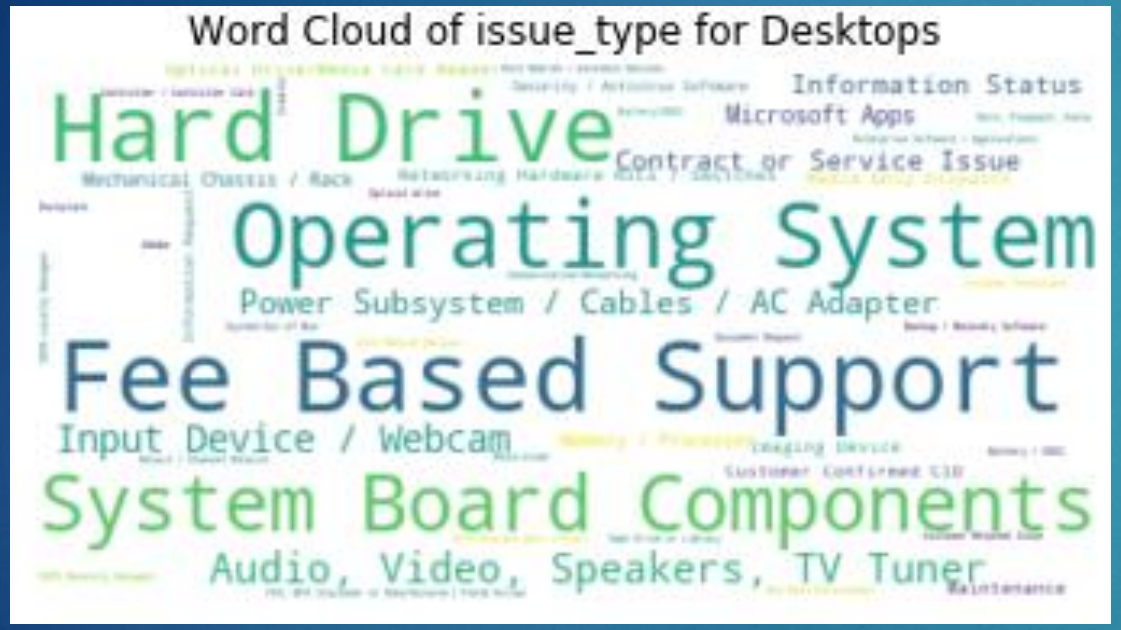
# Issue and Topic by Product Type (Laptop)

## Issues Identified by Customers

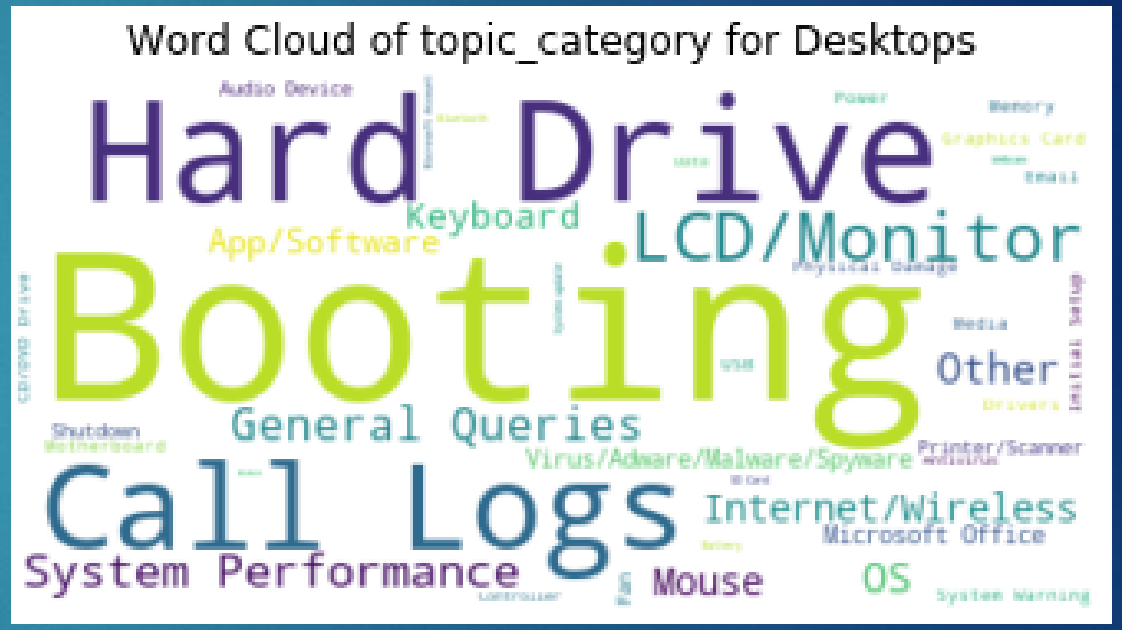

## Topics Identified by Agents

# Issue and Topic by Product Type (Desktop)

## Issues Identified by Customers

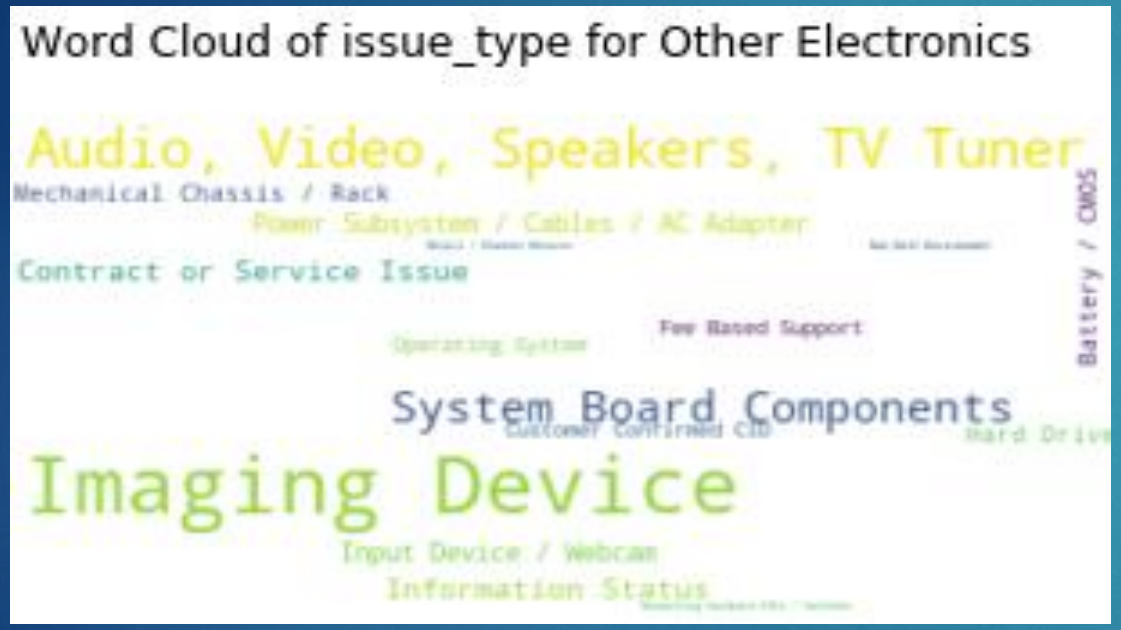



## Topics Identified by Agents

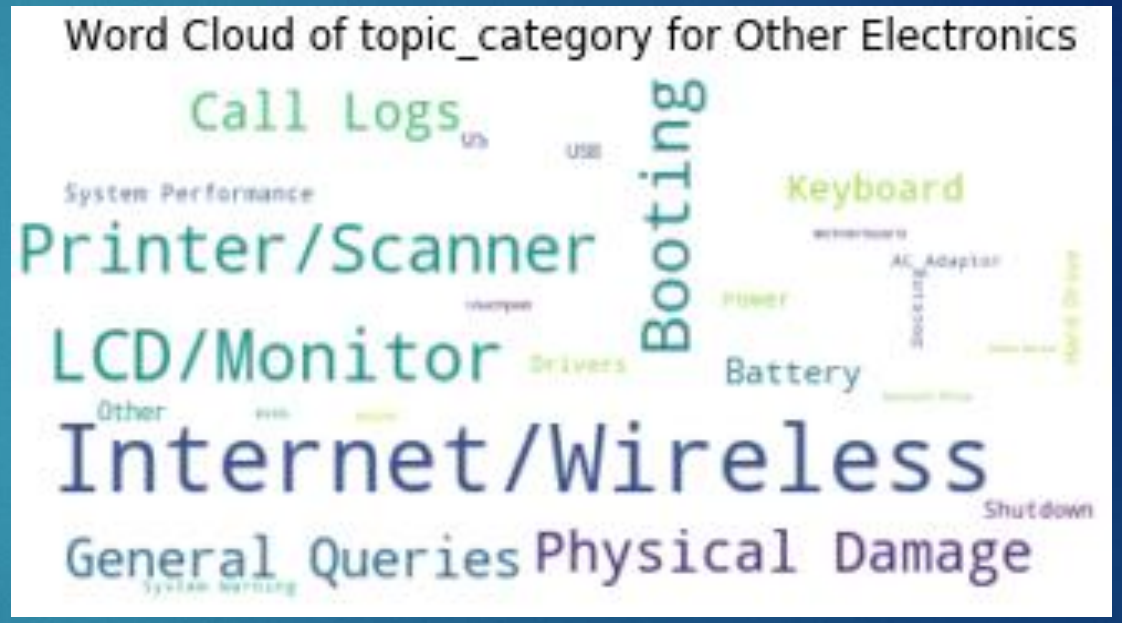


# Issue and Topic by Product Type (Other Electronics)

### Issues Identified by Customers

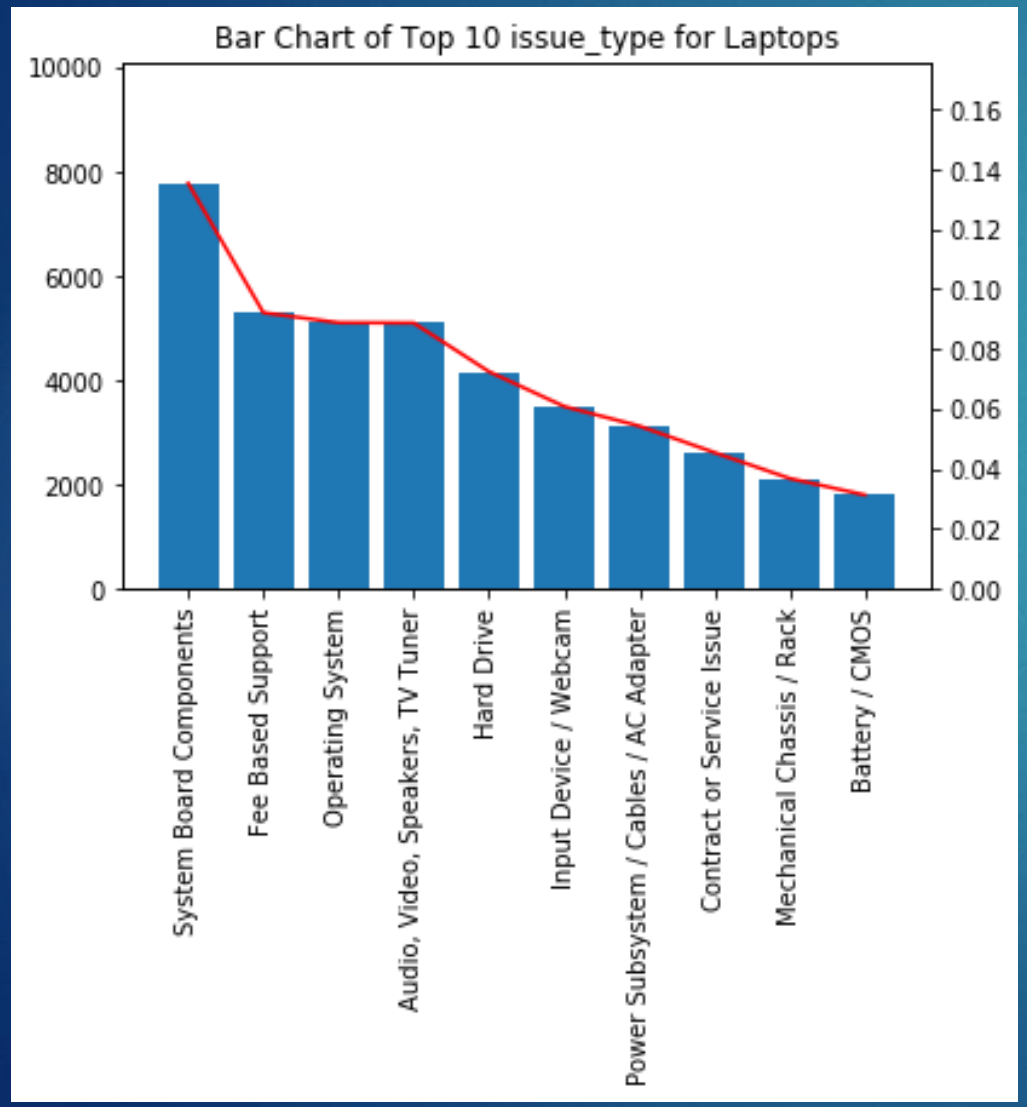



### Topics Identified by Agents

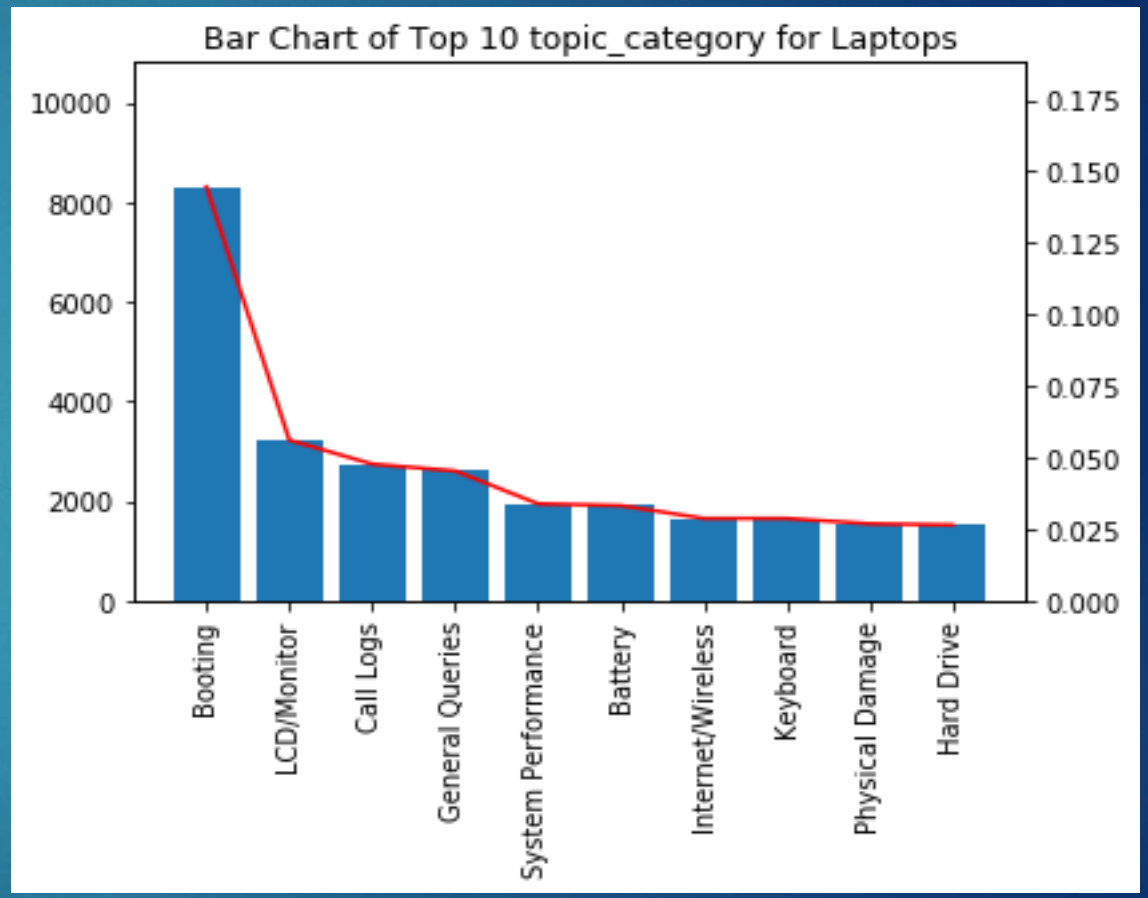


# Issue and Topic by Product Type (Laptop)

## Issues Identified by Customers



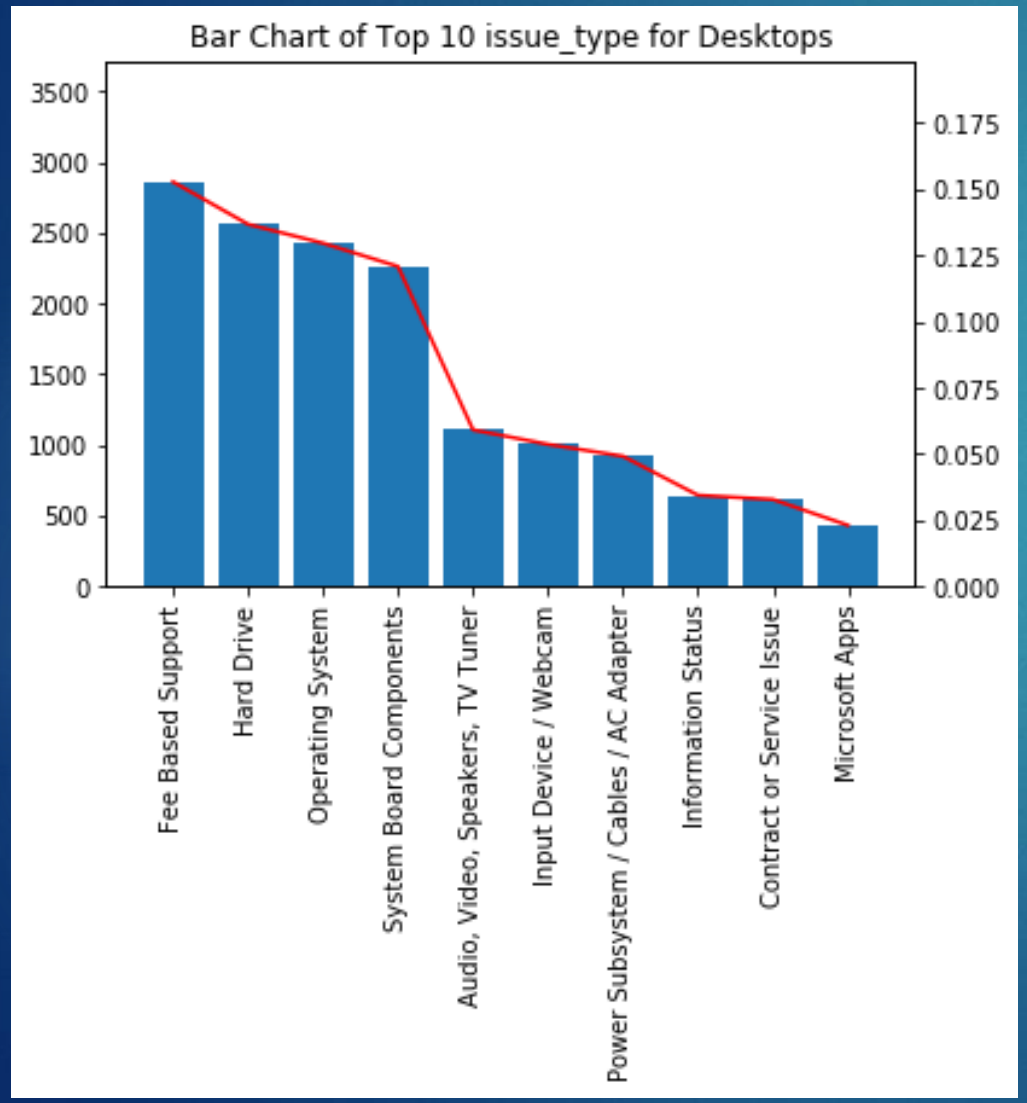

## Topics Identified by Agents



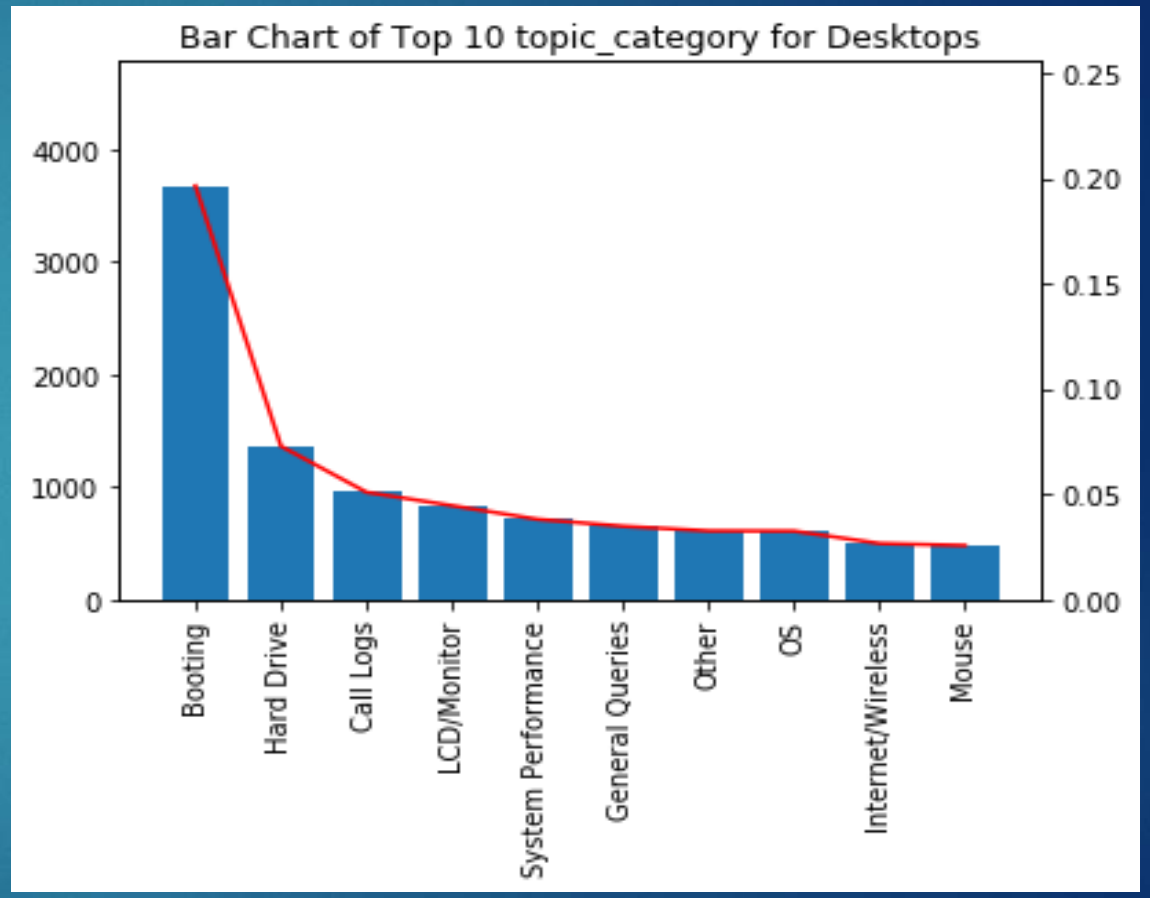

*Booting is always the biggest issue

# Issue and Topic by Product Type (Desktop)

## Issues Identified by Customers



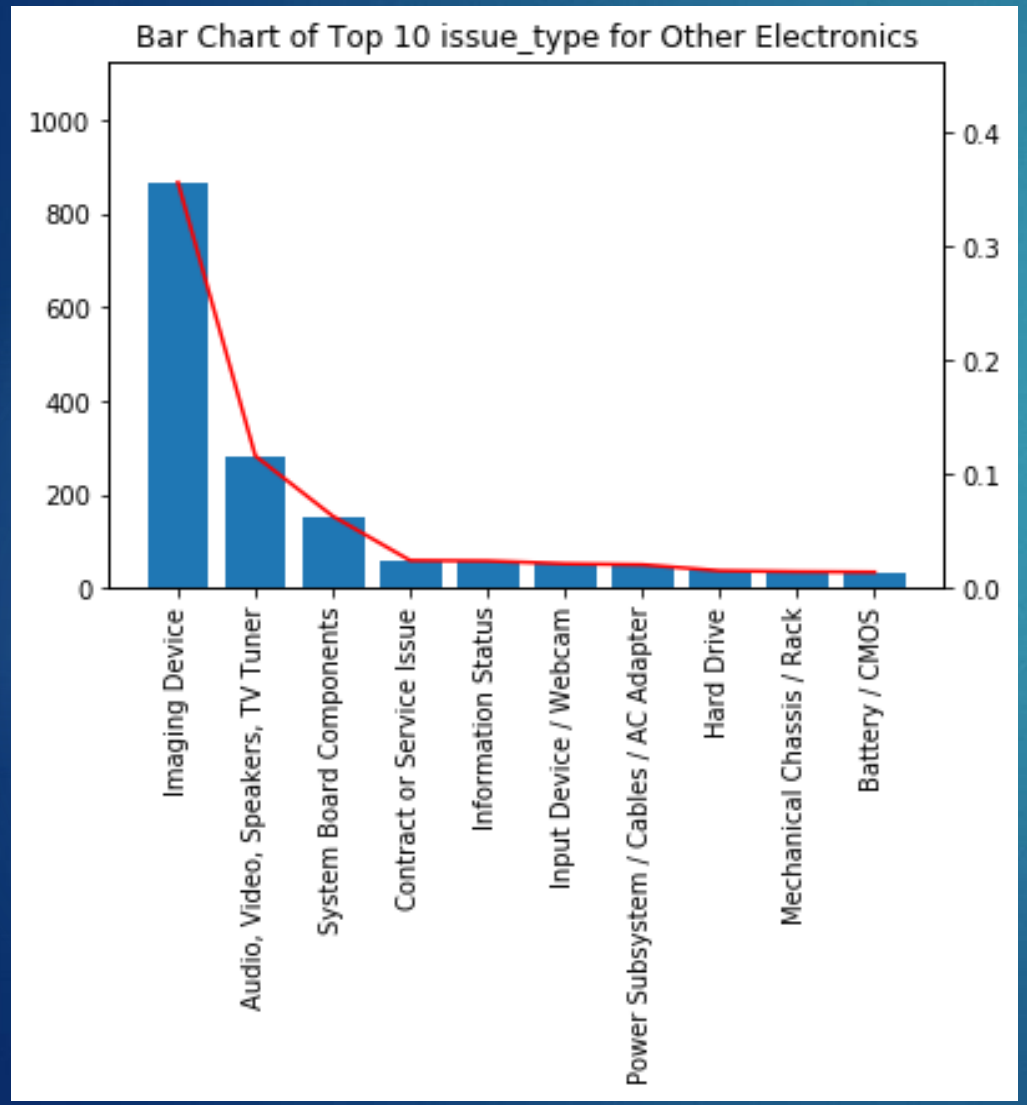

## Topics Identified by Agents



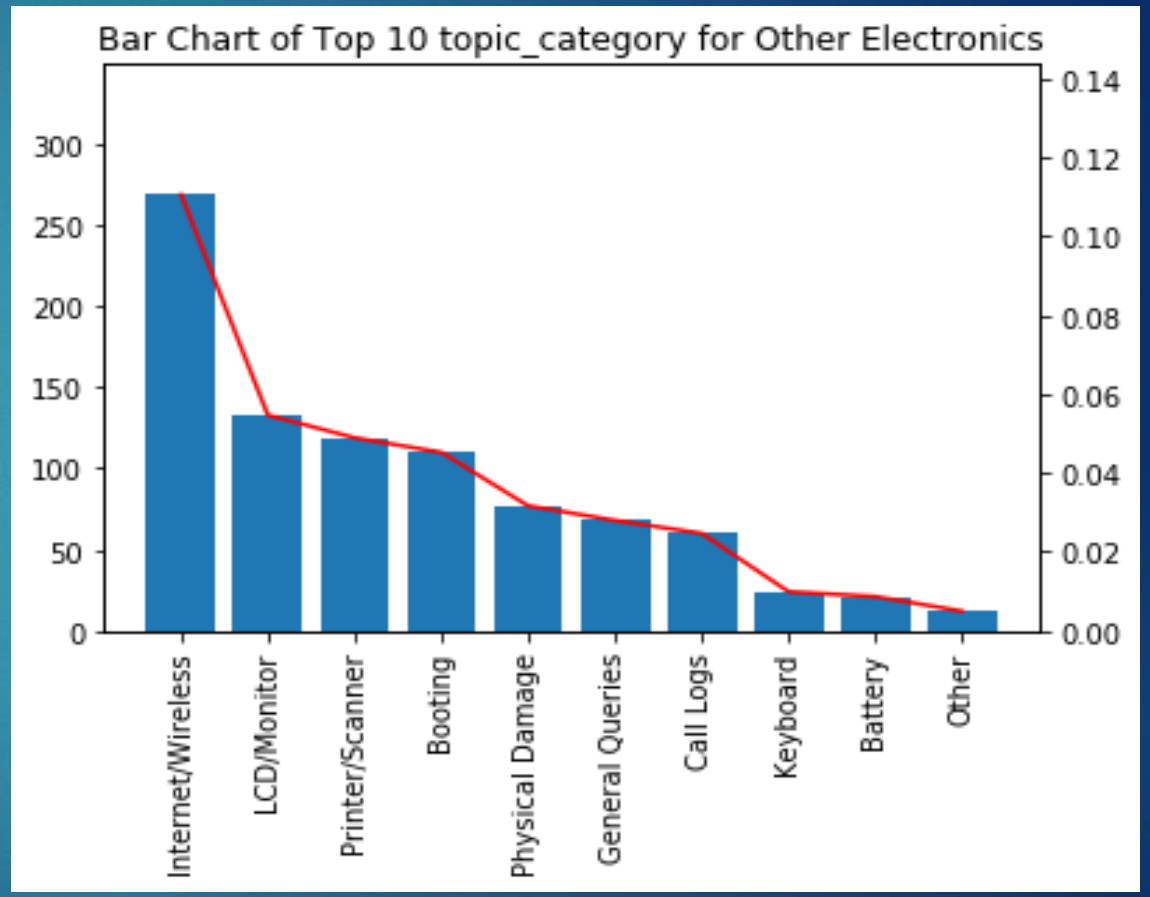

*Need to pay more attention to the second place of the list that is hard drive for desktops

# Issue and Topic by Product Type (Other Electronics)

## Issues Identified by Customers




## Topics Identified by Agents



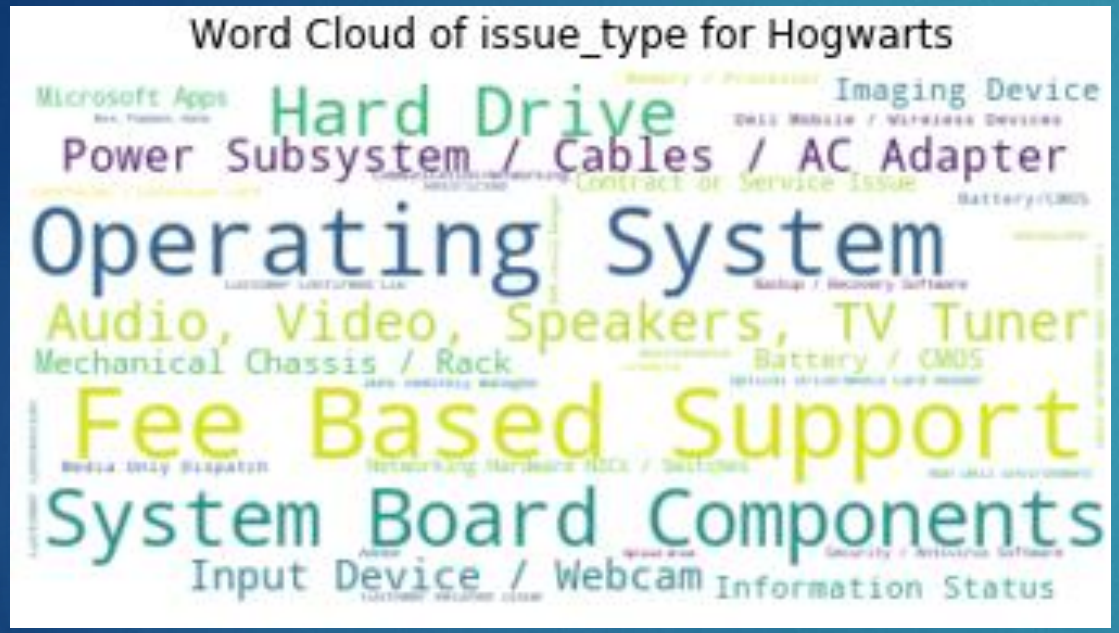

***Internet/Wireless issue turns out to dominate the other electronics cases**

# Issue and Topic by Region

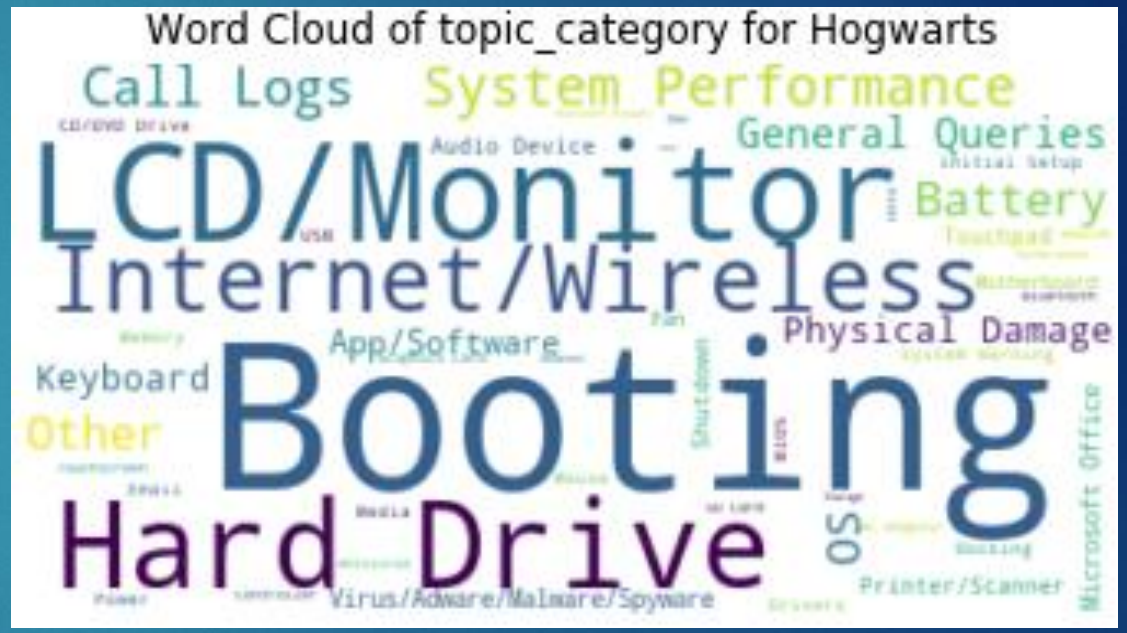
- Regions include:
  - Hogwarts
  - Middle Earth
  - Milky Way
- Items included:
  - Word Cloud
  - Bar Charts including both counts and proportions

# Issue and Topic by Region (Hogwarts)

### Issues Identified by Customers

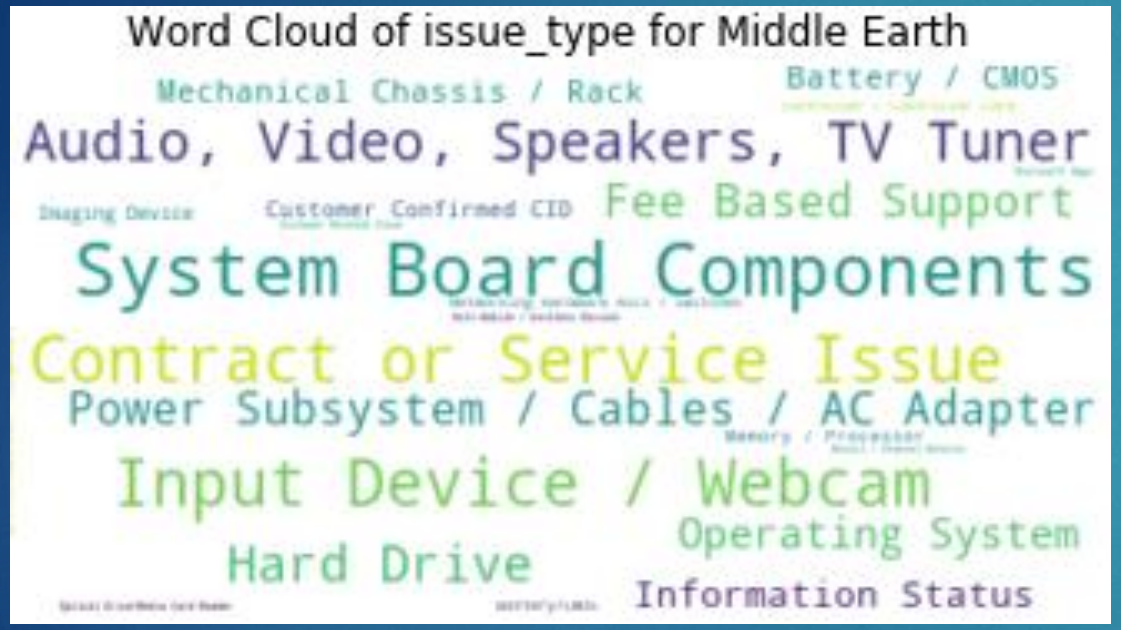



### Topics Identified by Agents

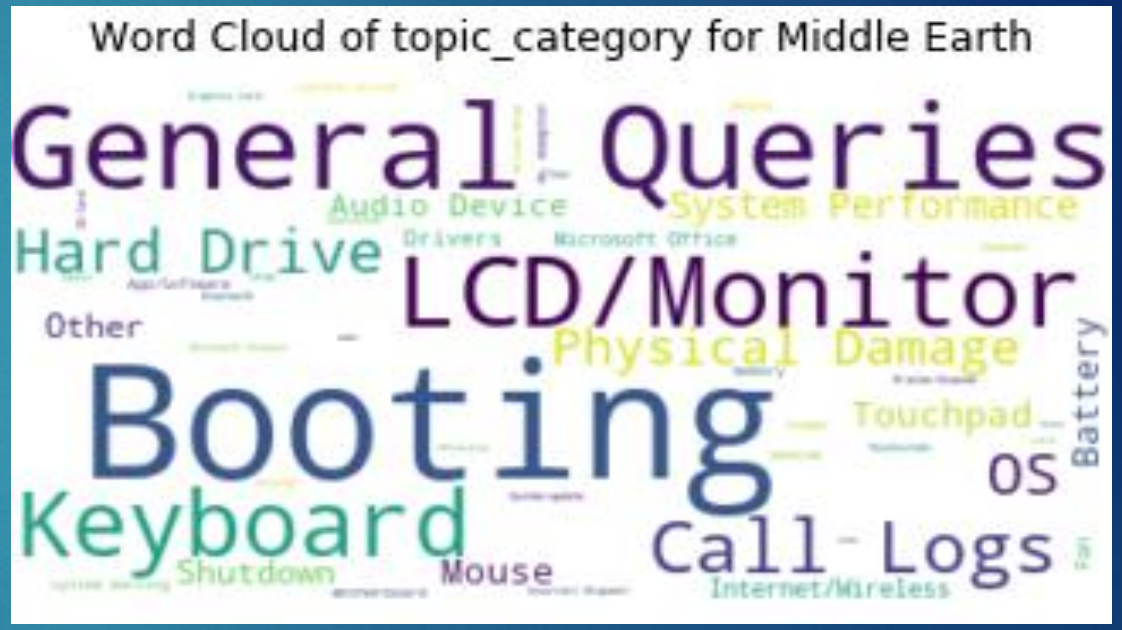


# Issue and Topic by Region (Middle Earth)

## Issues Identified by Customers




## Topics Identified by Agents



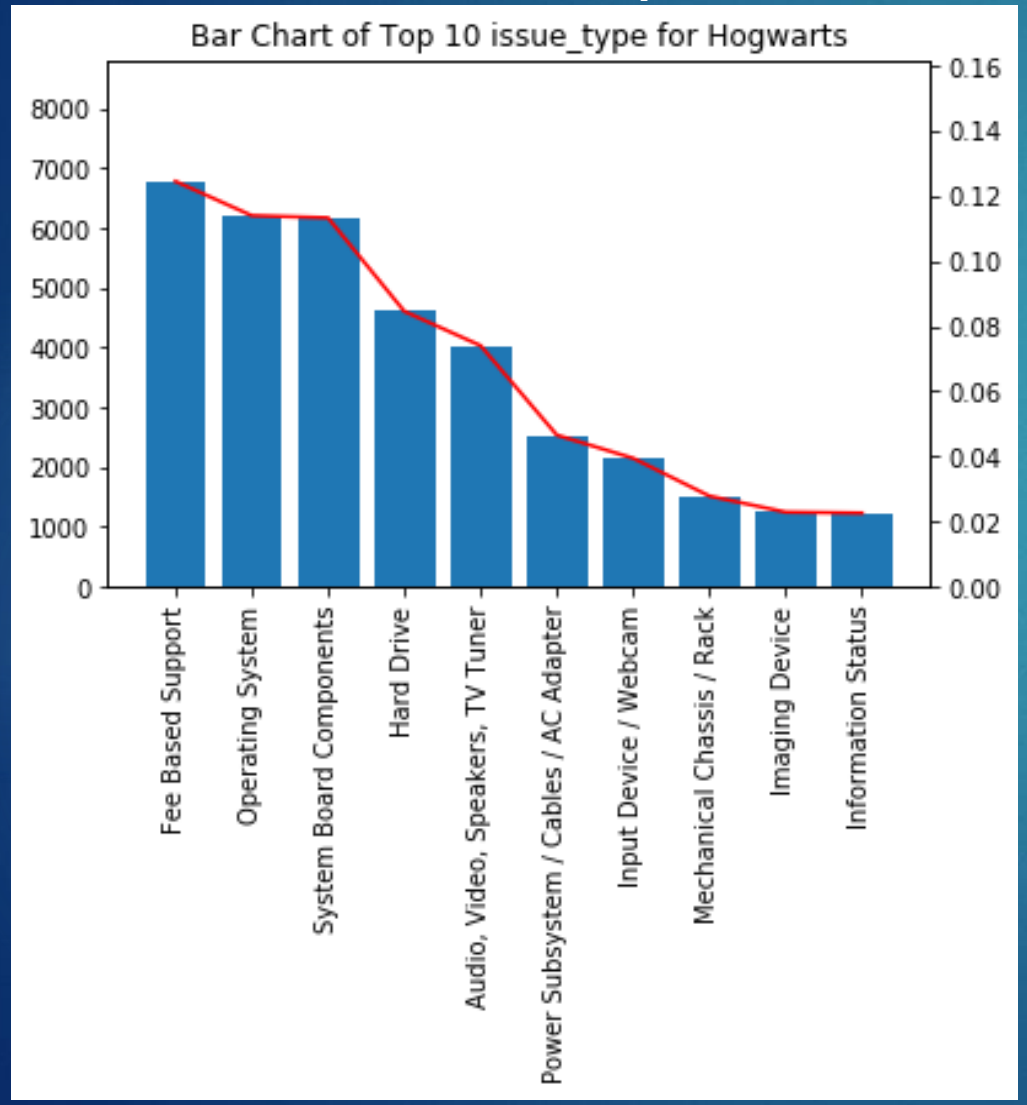
# Issue and Topic by Region (Milky Way)

### Issues Identified by Customers

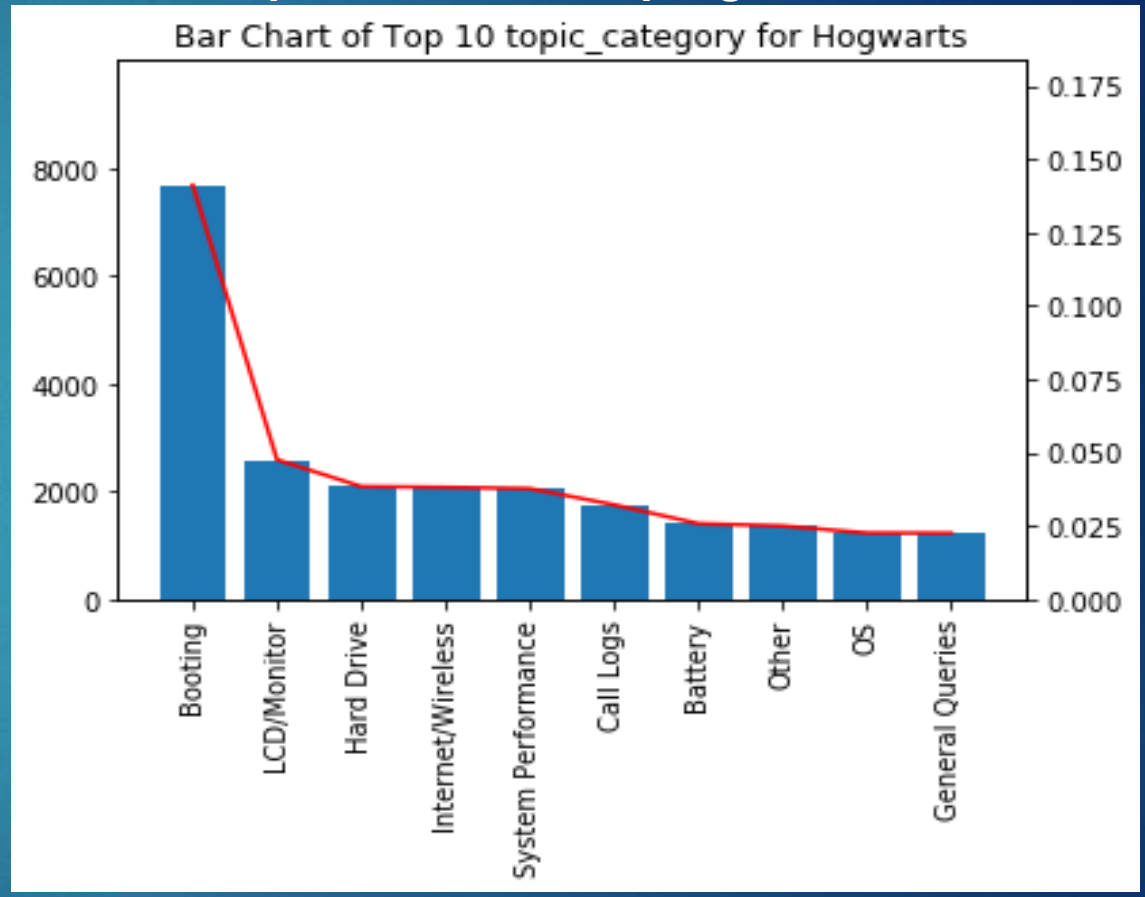

### Topics Identified by Agents

# Issue and Topic by Region (Hogwarts)

## Issues Identified by Customers

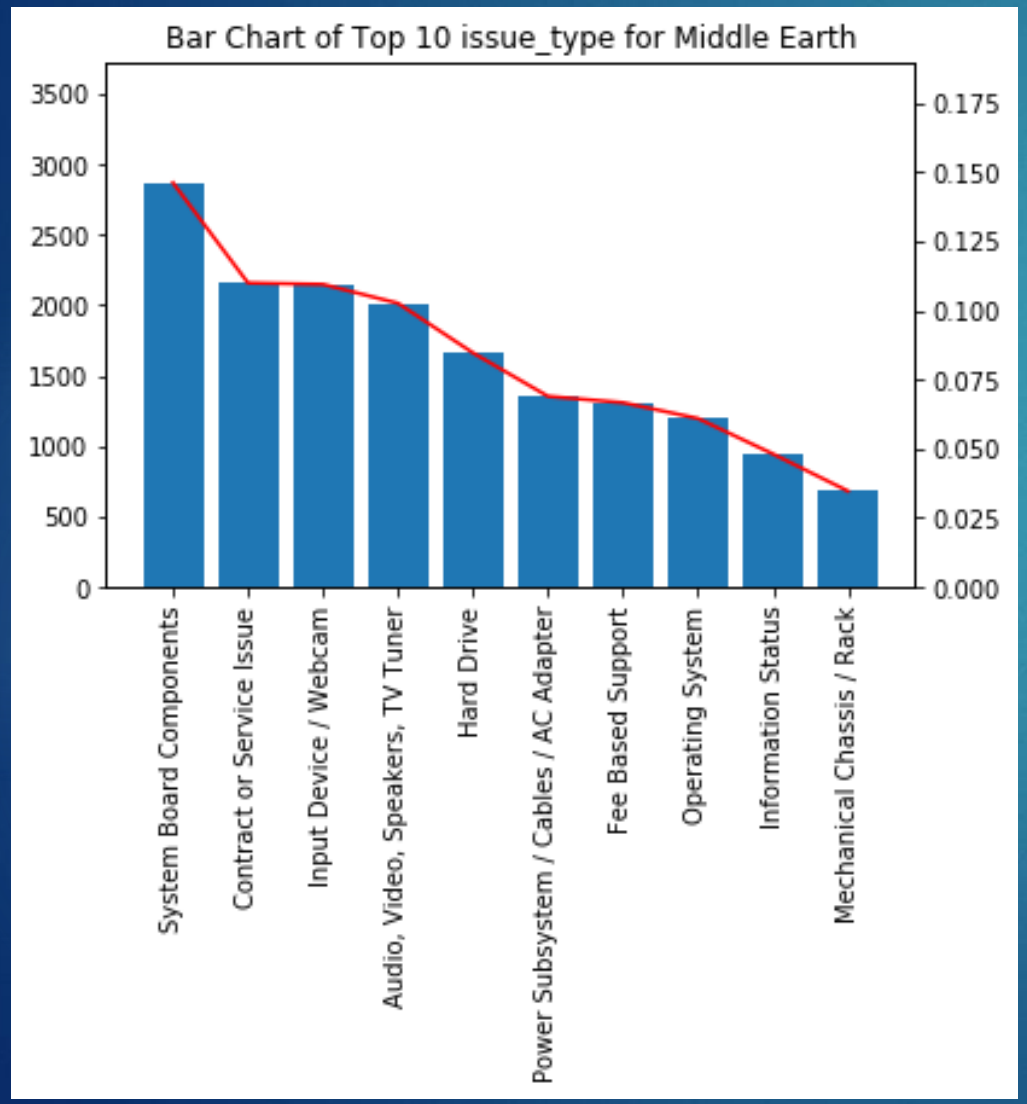



## Topics Identified by Agents

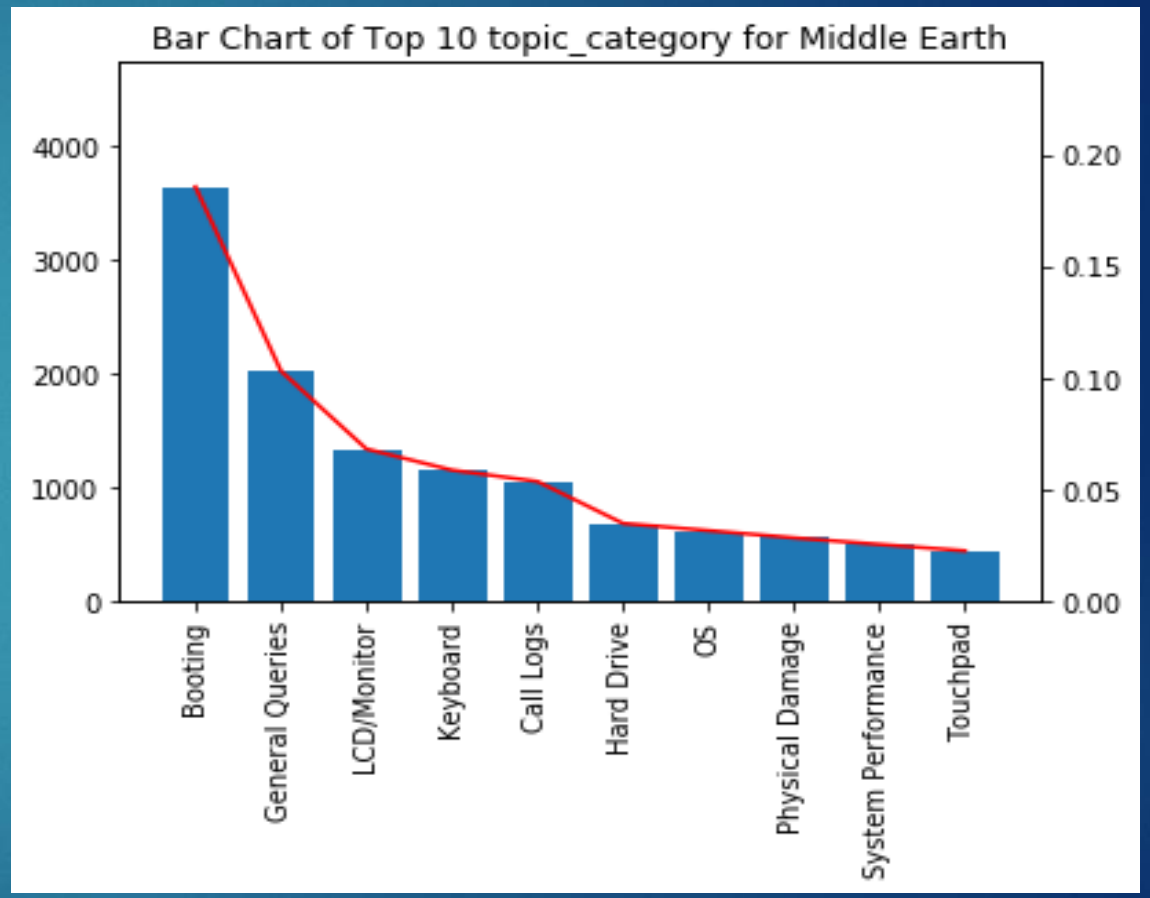


# Issue and Topic by Region (Middle Earth)

## Issues Identified by Customers

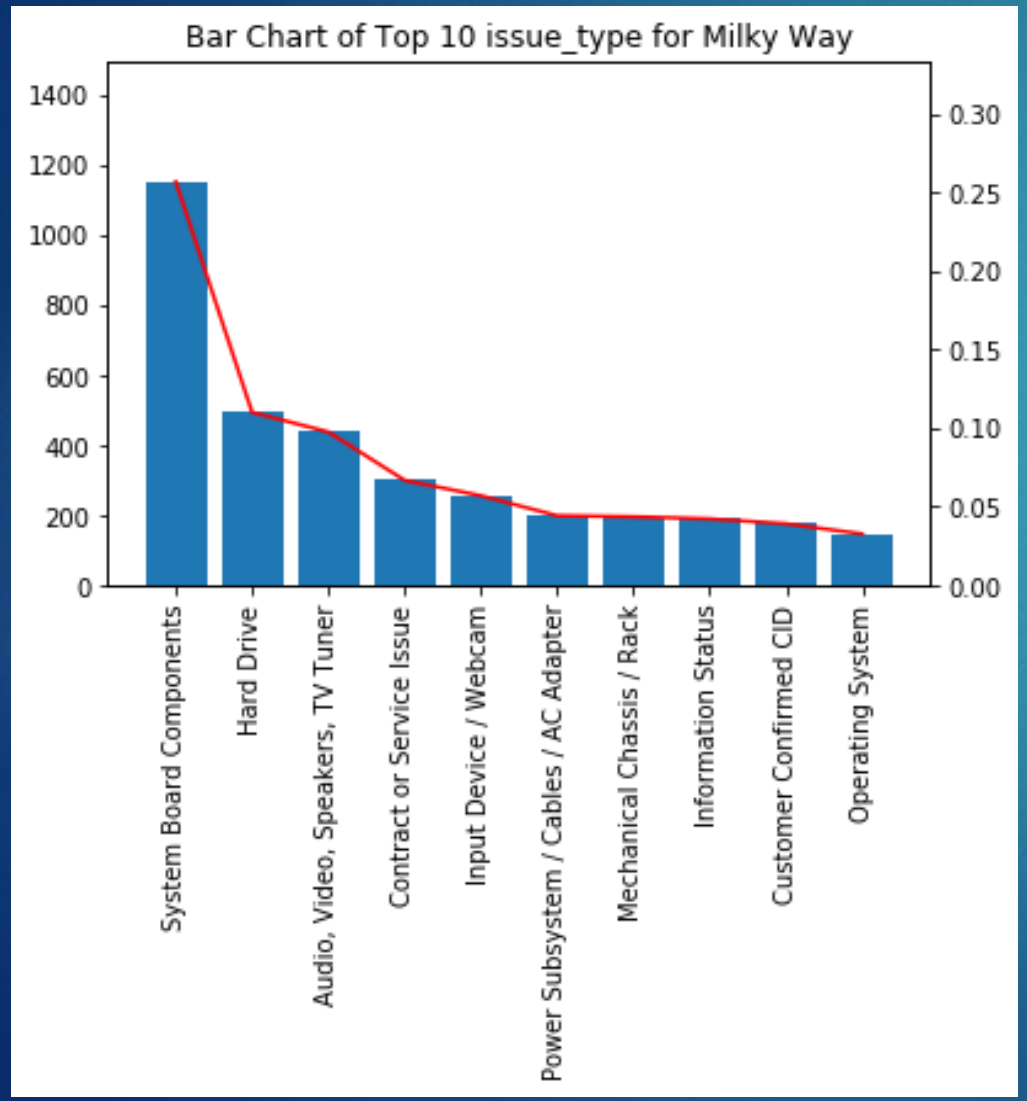



## Topics Identified by Agents

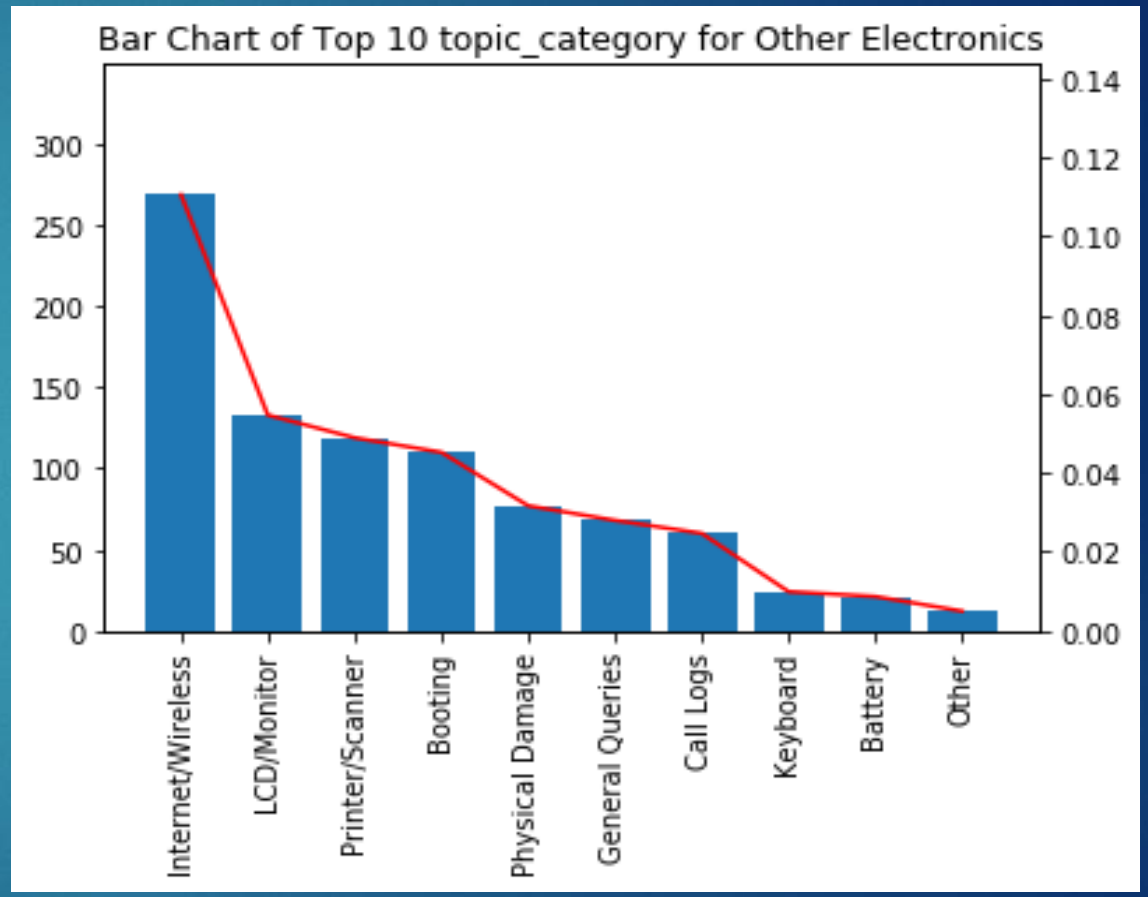


# Issue and Topic by Region (Milky Way)

## Issues Identified by Customers



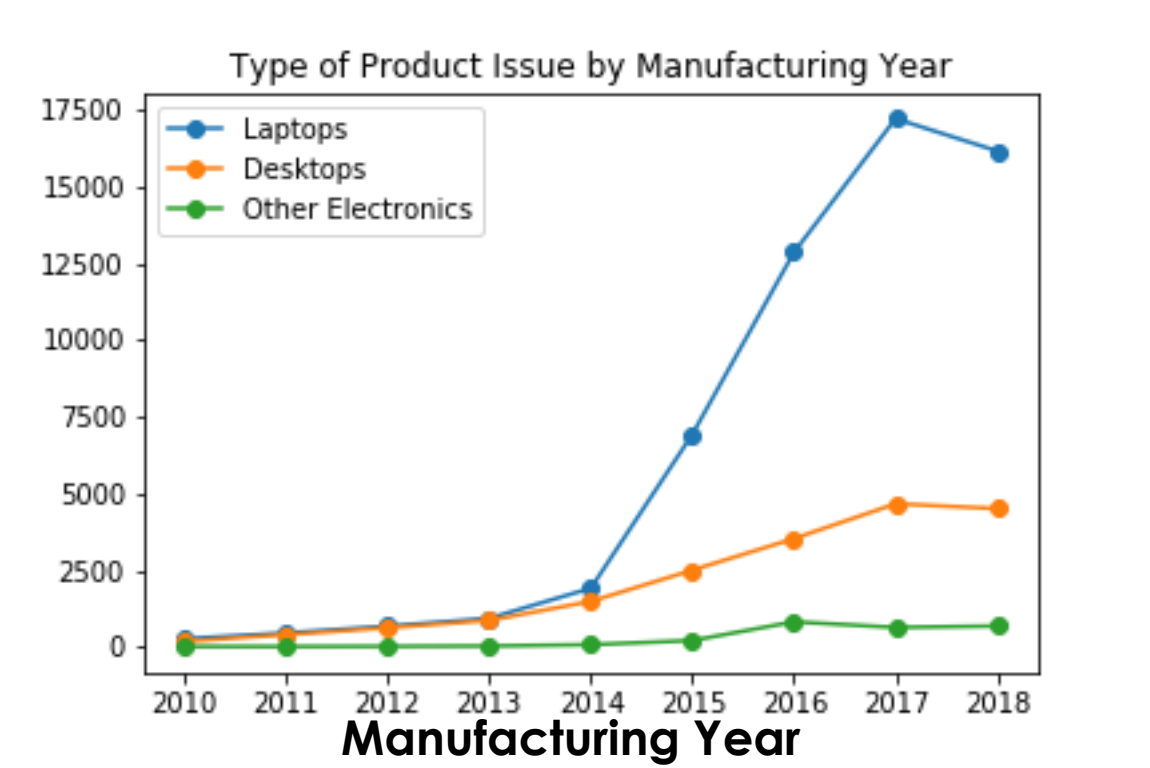

## Topics Identified by Agents




*while booting, LCD and hard drive are usually the main issue, internet/wireless had a bigger portion of issue in Milky Way.

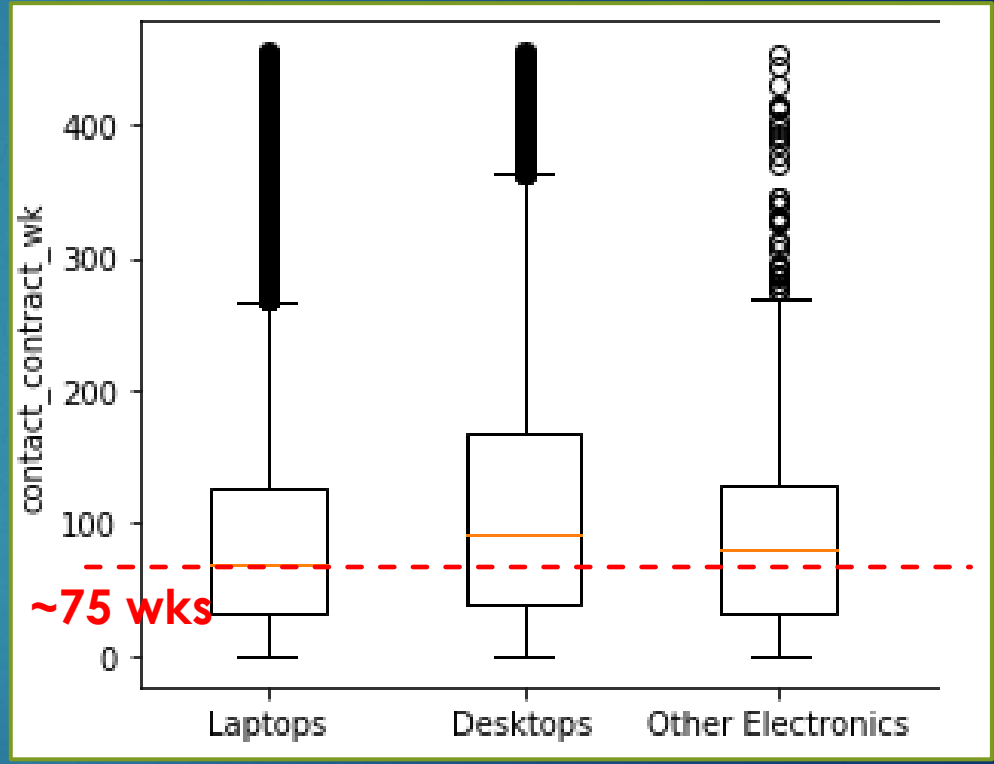
# Year Distribution and Time for Product to start having Issues

## Number of Instances for Product Types by Manufacturing Year




*Even though more instances were spotted after 2013, but it also implied sales were also increasing. Need to have the sales number for each category for the proportion number.

## How soon customer contacted after the warranty started




** Most of contacts came in around 75 weeks after warranty started. Desktop seems to have more robust quality but more variations than the others

# Services/Fixing

- For the **service/fixing analysis**, several aspects were also explored:
  - Top issues and topics reported by different contact type
  - Top issues and topics handled by different repair types (Hard of Soft)
  - Top issues involving managers
  - Parts:
    - Top sent/resent parts to be replaced for the issues/fixing
    - Year distribution for the top sent/resent parts and **how soon** needing replacement after warranty started
  - Something about the agents: Diagnostic usages, number of sent parts and contacting manager flag vs. agent tenure in days

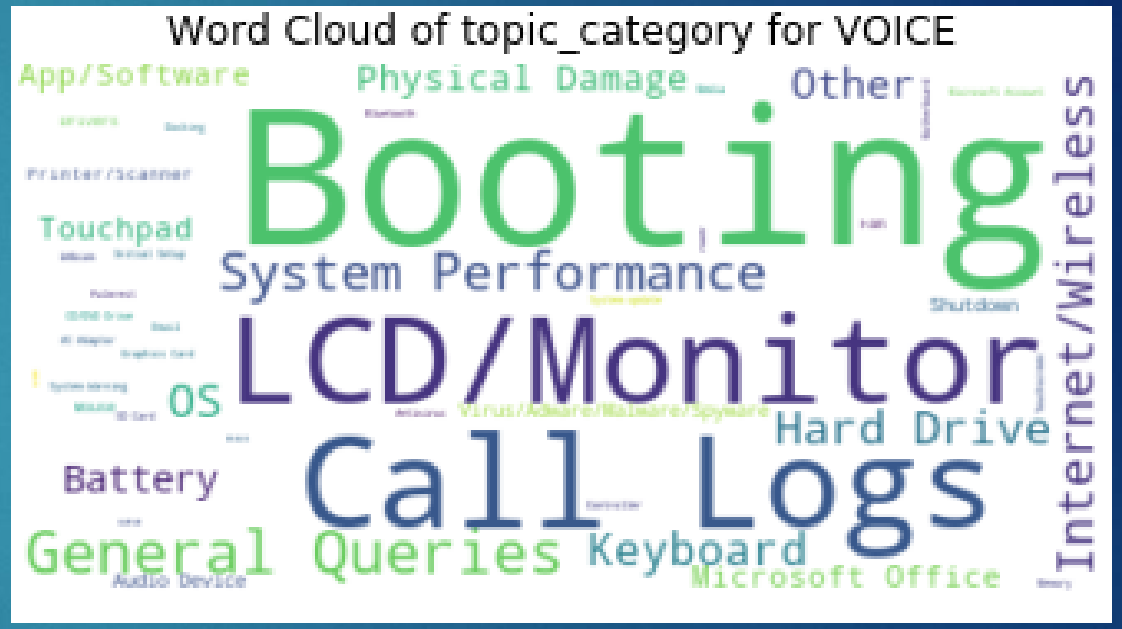
# Issue and Topic by Contact Type

- Contact Type:
  - Voice
  - Chart
  - Email
- Items included:
  - Word Cloud
  - Bar Charts including both counts and proportions

# Issue and Topic by Contact Type (Voice)

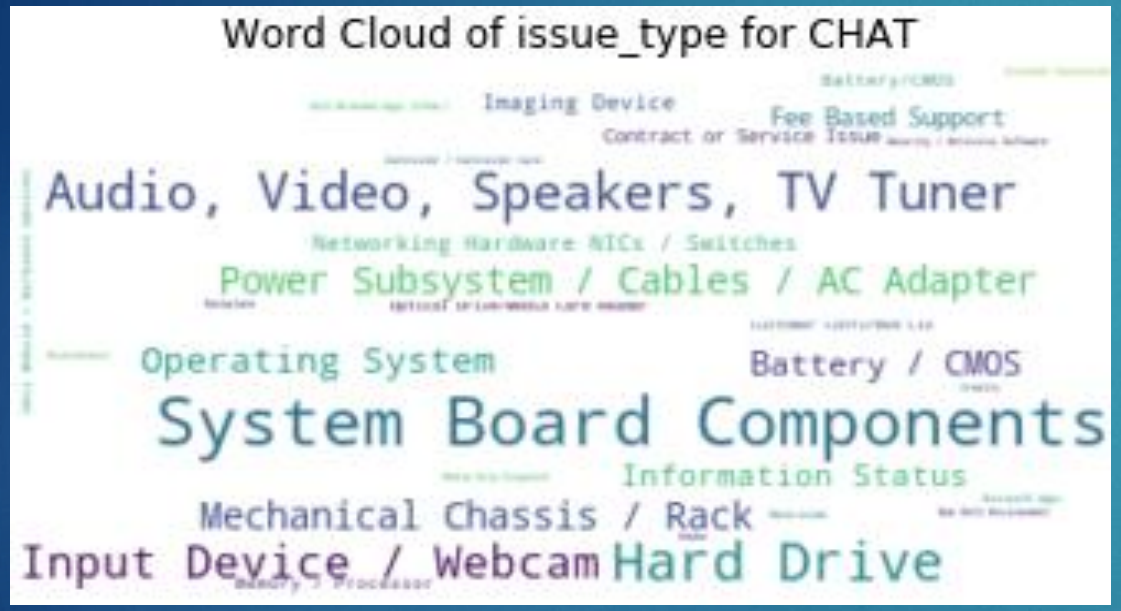
## Issues Identified by Customers


## Topics Identified by Agents



# Issue and Topic by Contact Type (Chat)

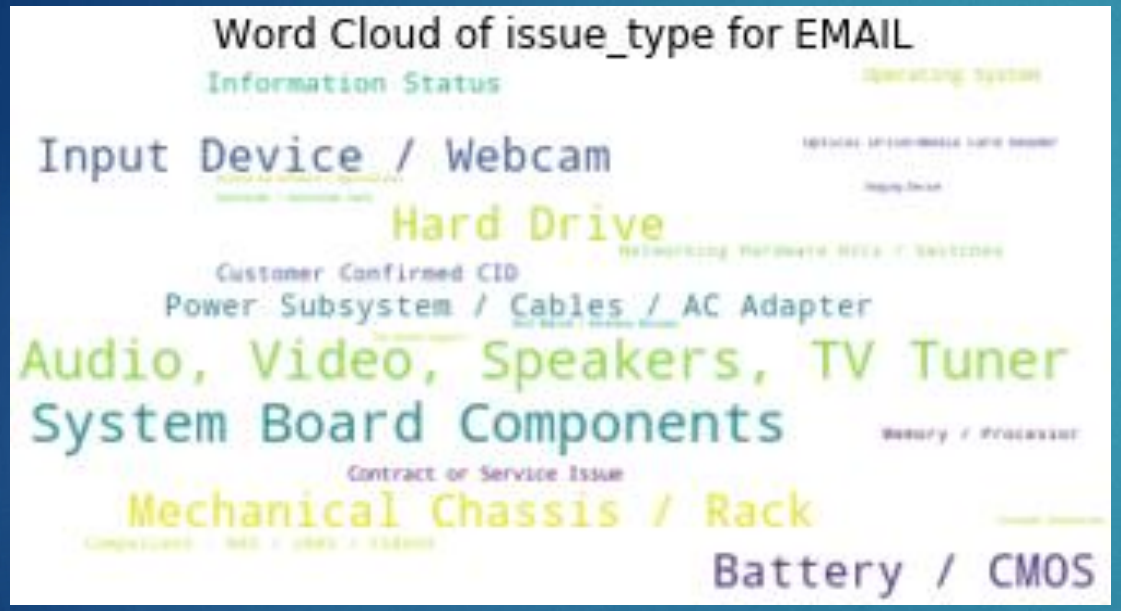
### Issues Identified by Customers



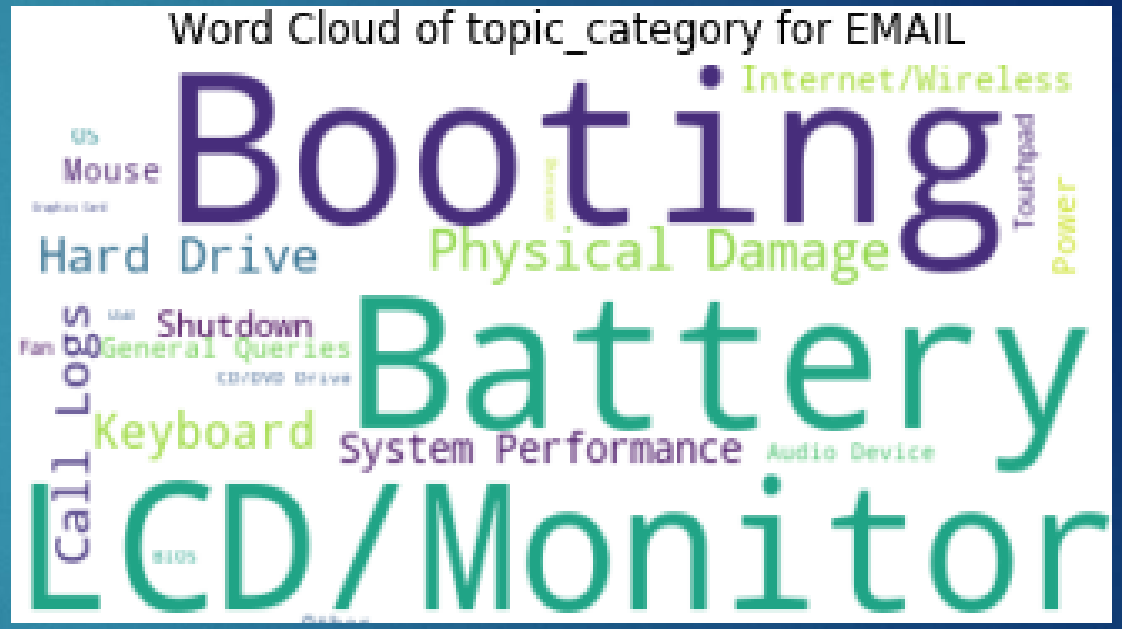

### Topics Identified by Agents

# Issue and Topic by Contact Type (Email)

### Issues Identified by Customers

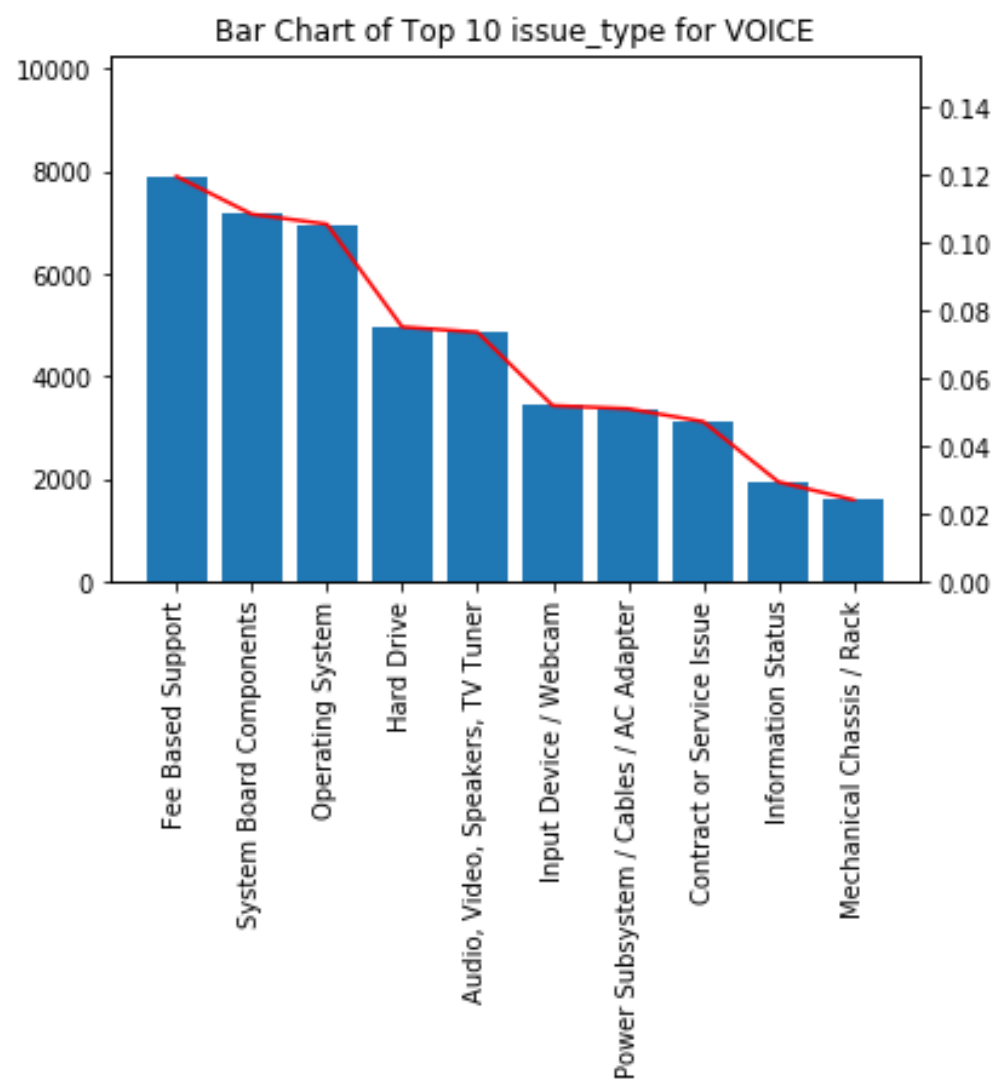



### Topics Identified by Agents

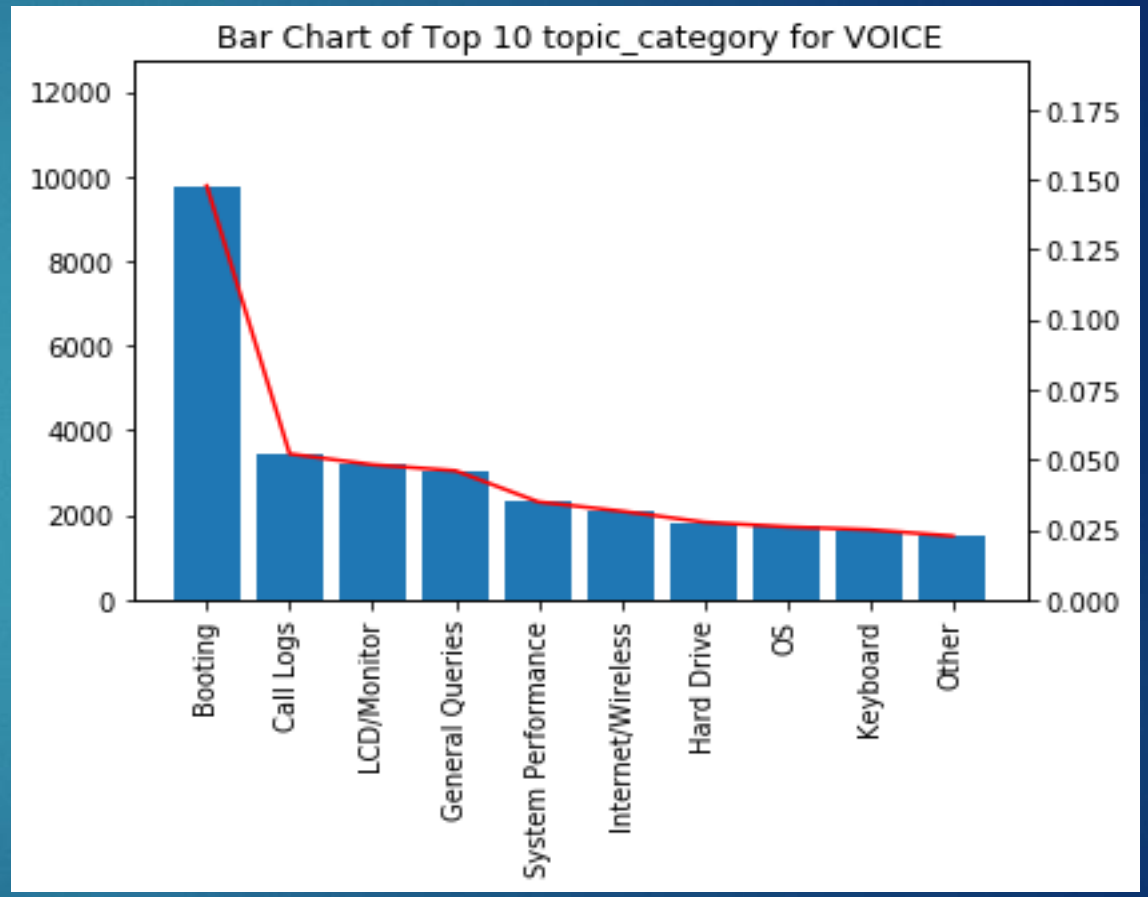


# Issue and Topic by Contact Type (Voice)

## Issues Identified by Customers

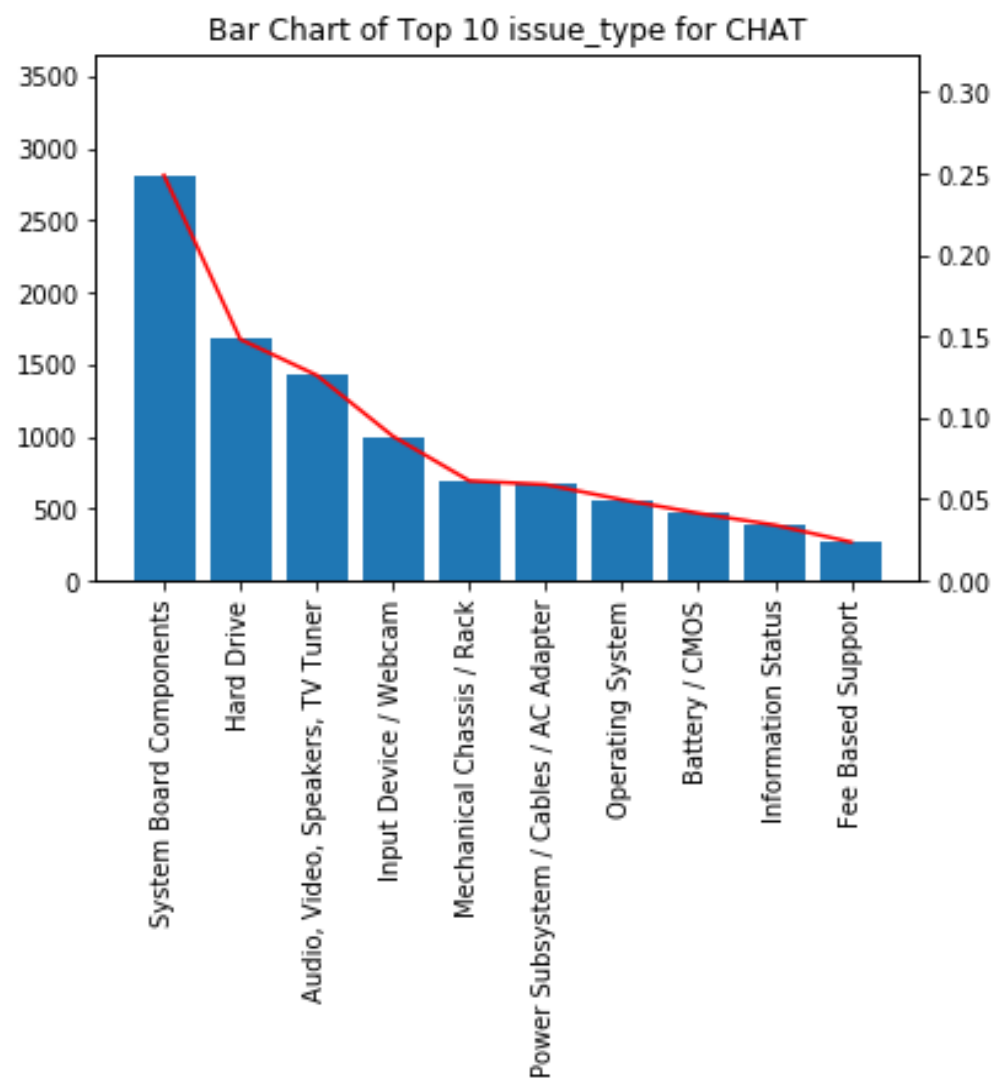



## Topics Identified by Agents

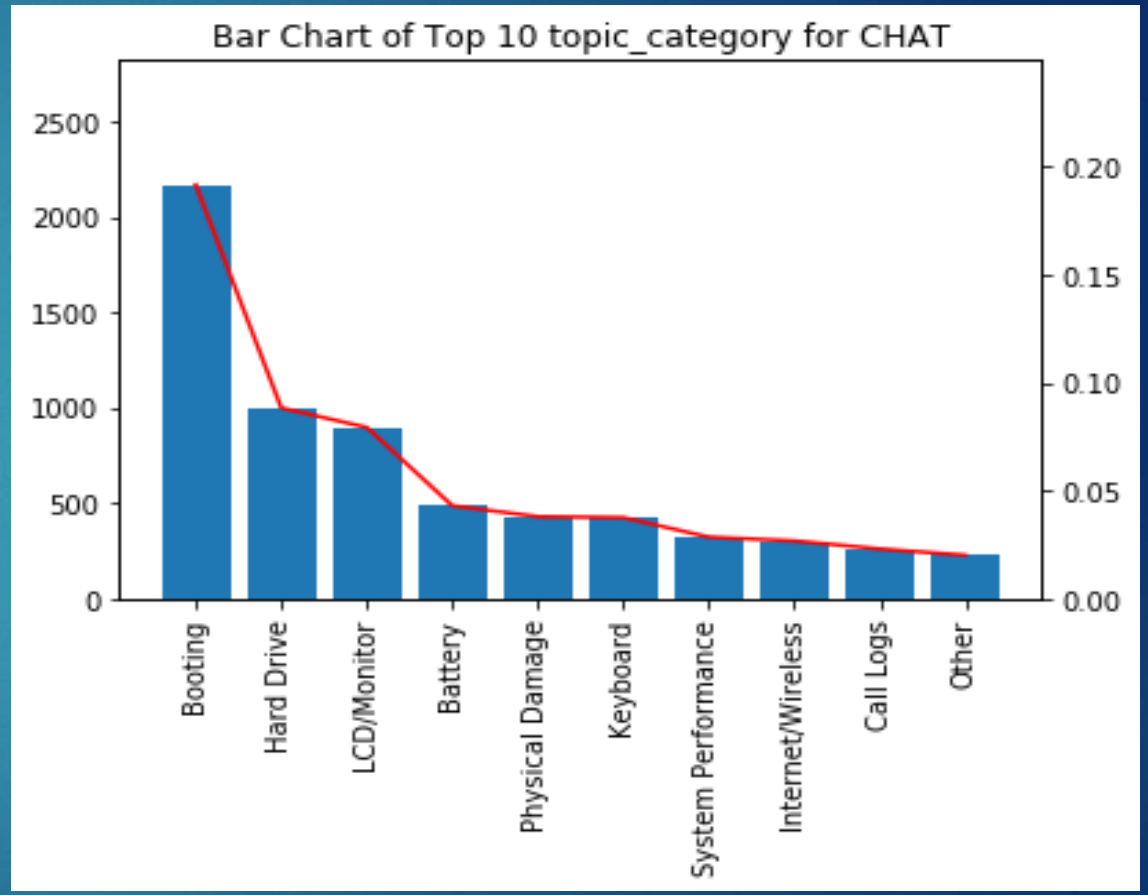


# Issue and Topic by Contact Type (Chat)

## Issues Identified by Customers

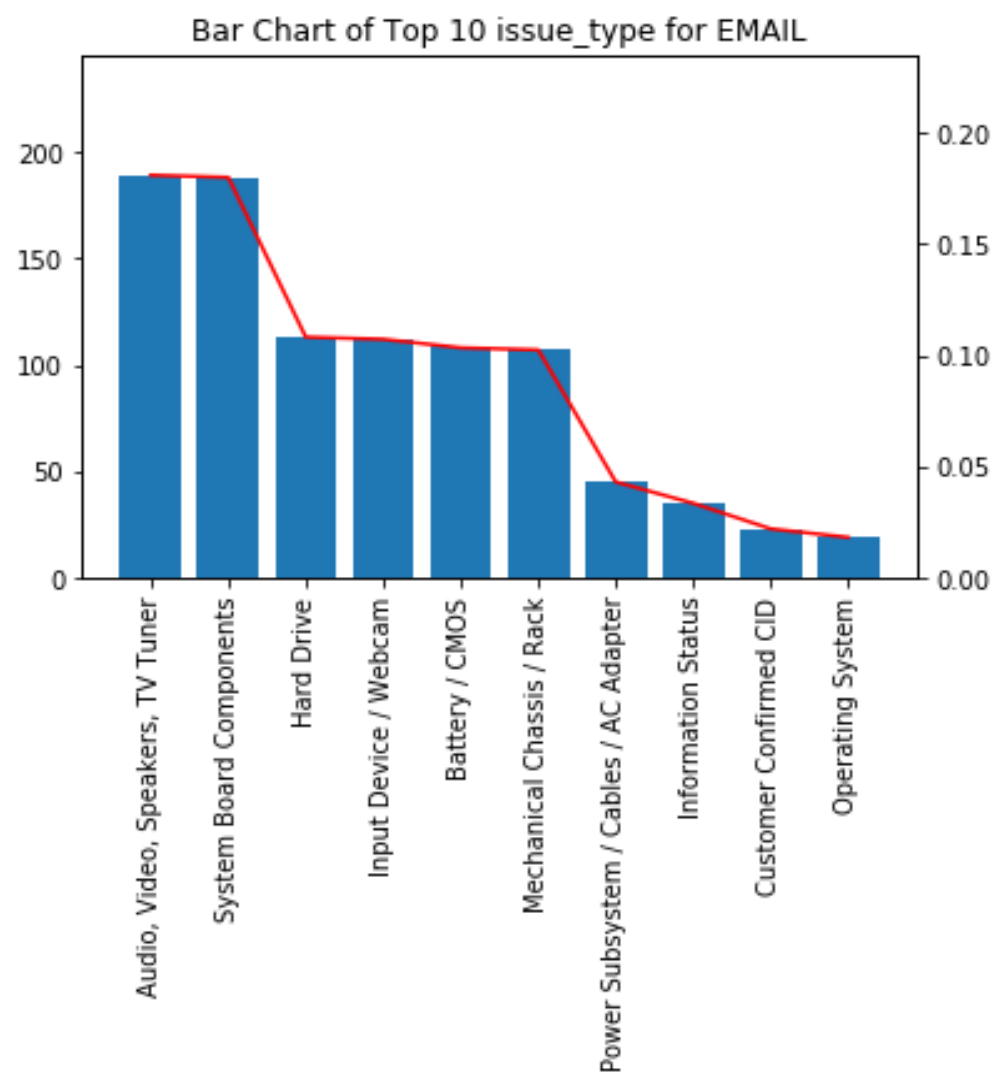



## Topics Identified by Agents

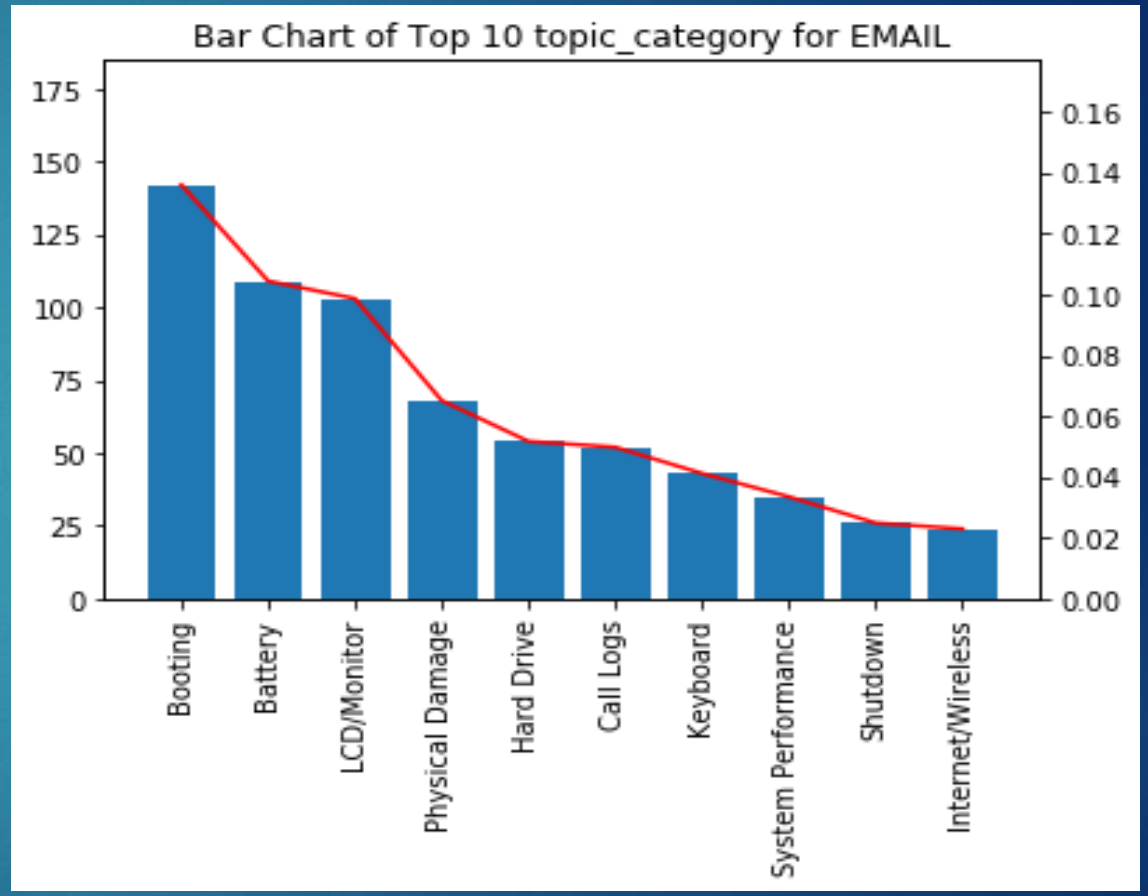


# Issue and Topic by Contact Type (Email)

## Issues Identified by Customers




## Topics Identified by Agents



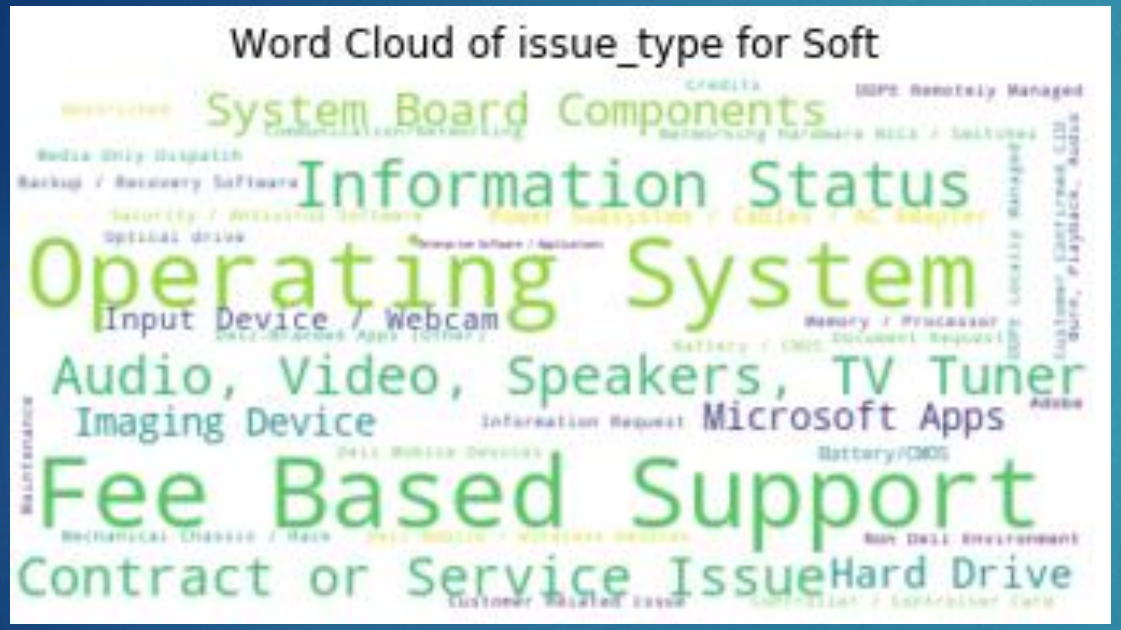

*The main issues are booting and LCD for all three contact types

# Issue and Topic by Repair Type

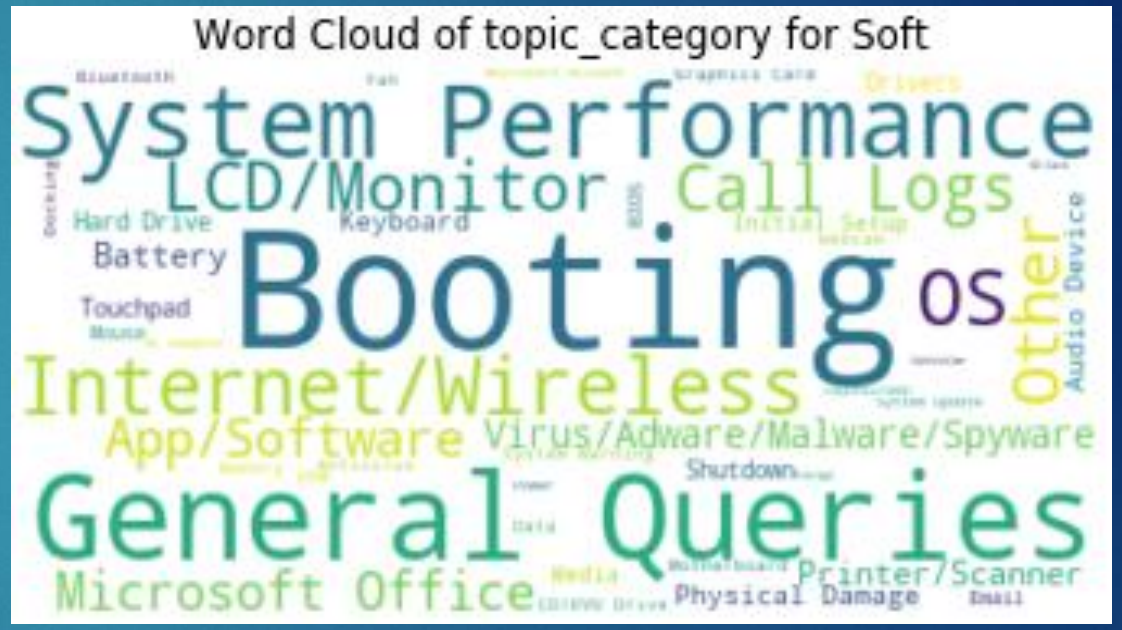
- Repair Type: Soft and Hard
- Items included:
  - Word Cloud
  - Bar Charts including both counts and proportions

# Issue and Topic Solved by Repair Type (Soft)

### Issues Identified by Customers




### Topics Identified by Agents



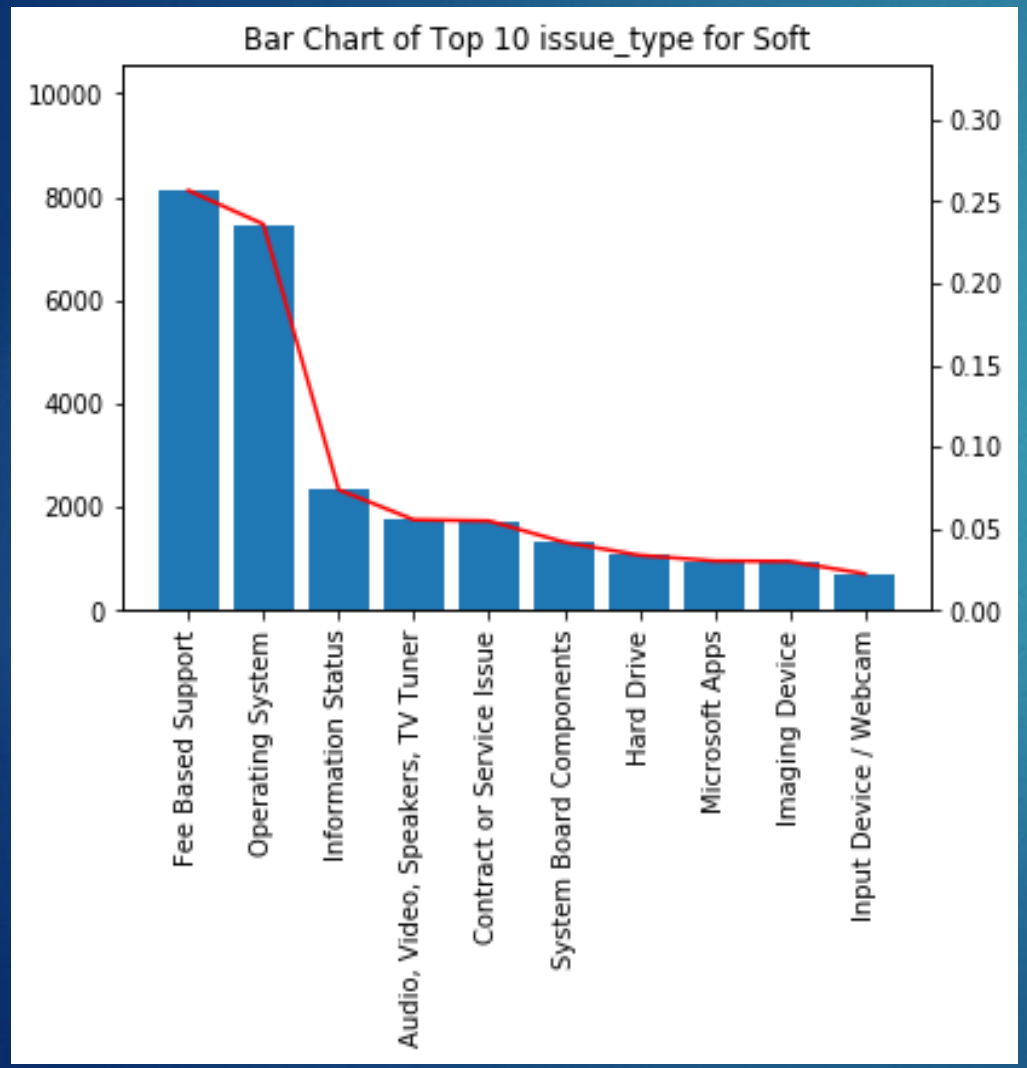
# Issue and Topic Solved by Repair Type (Hard)

**Issues Identified by Customers**

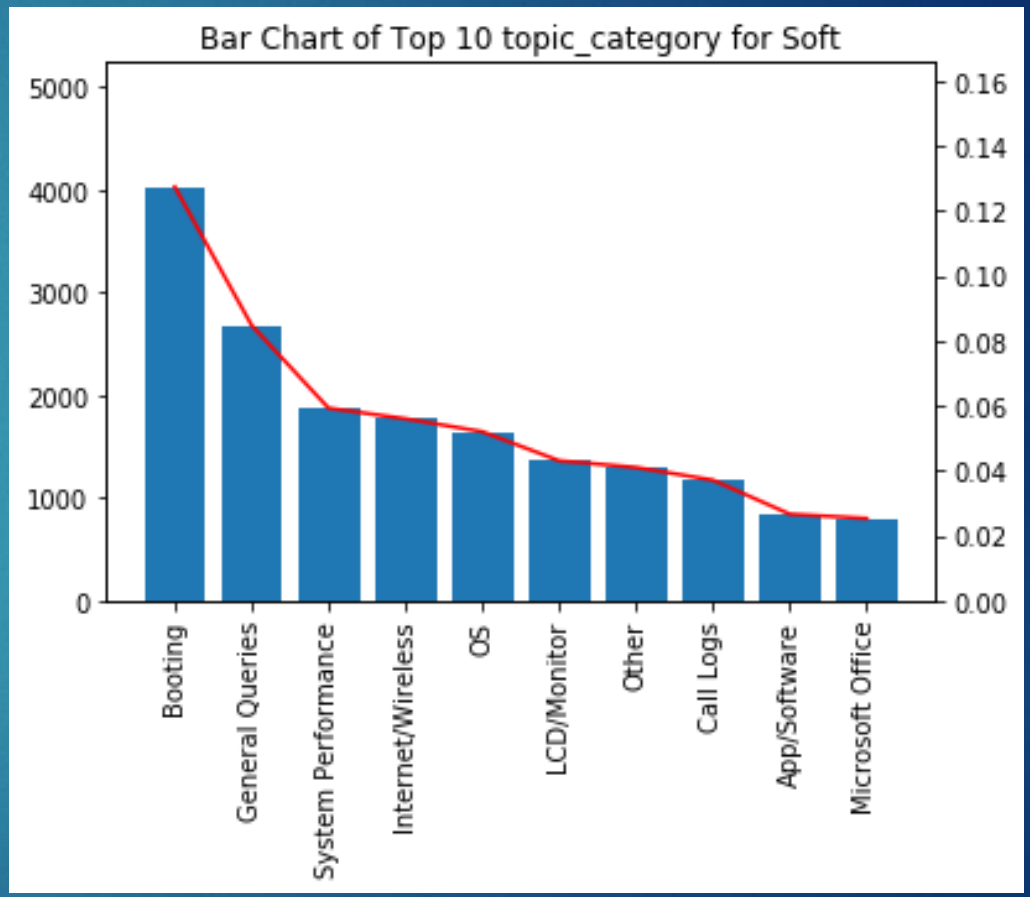

**Topics Identified by Agents**

# Issue and Topic Solved by Product Type (Soft)

## Issues Identified by Customers

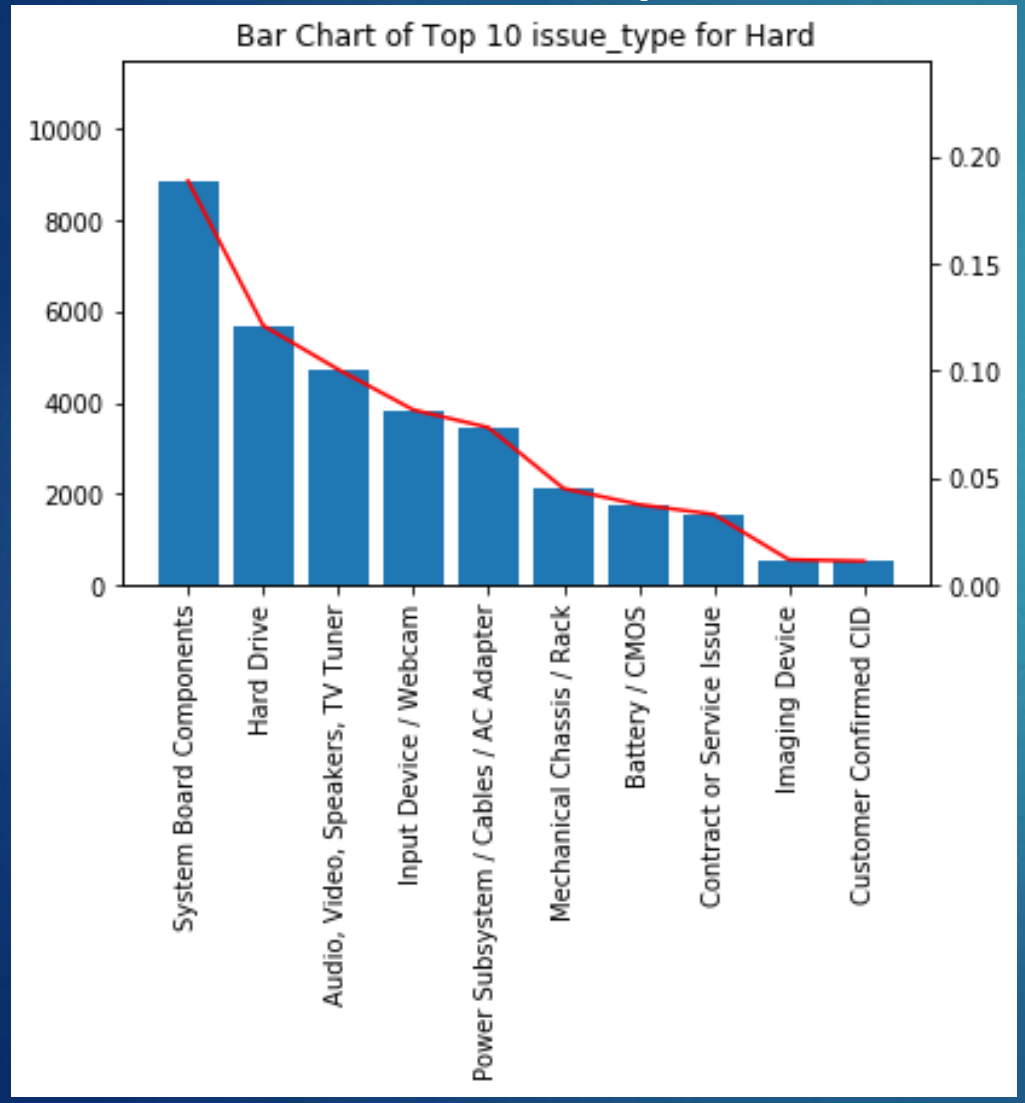



## Topics Identified by Agents

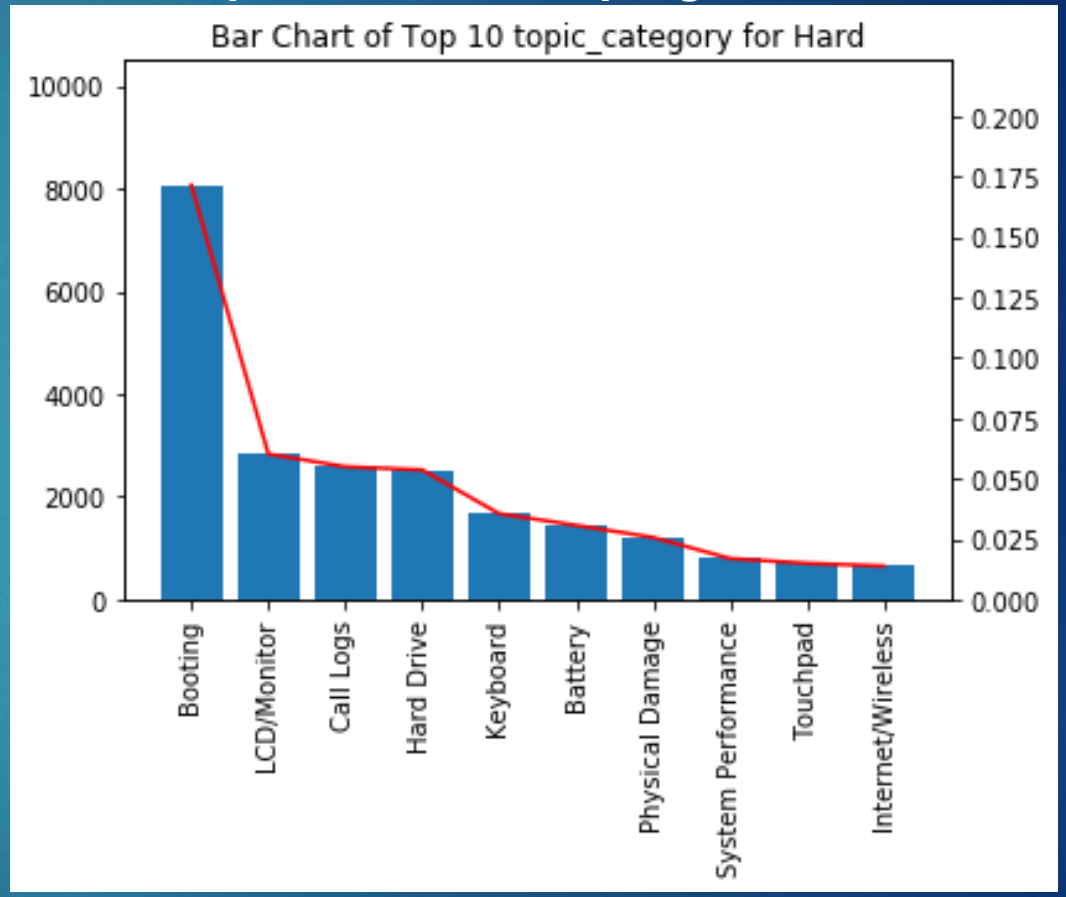


# Issue and Topic Solved by Product Type (Hard)

## Issues Identified by Customers



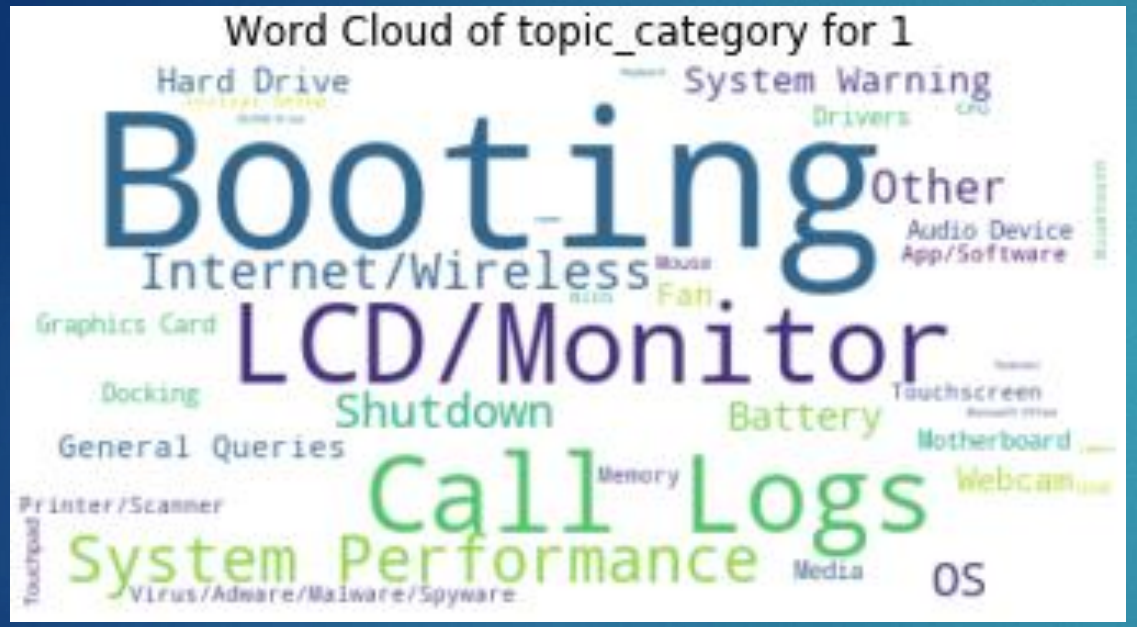

## Topics Identified by Agents

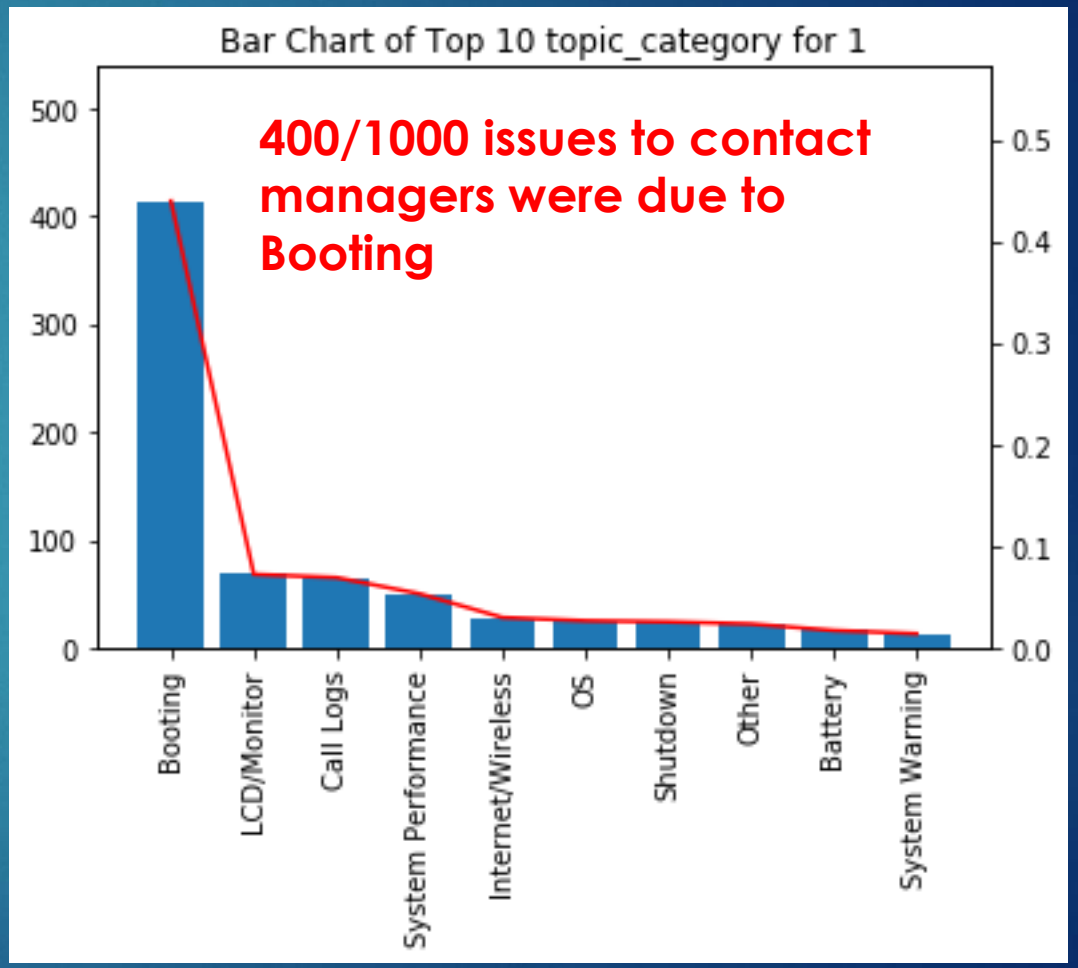



***Booting, LCD, Hard Drive and Battery requires more hard repairs.**

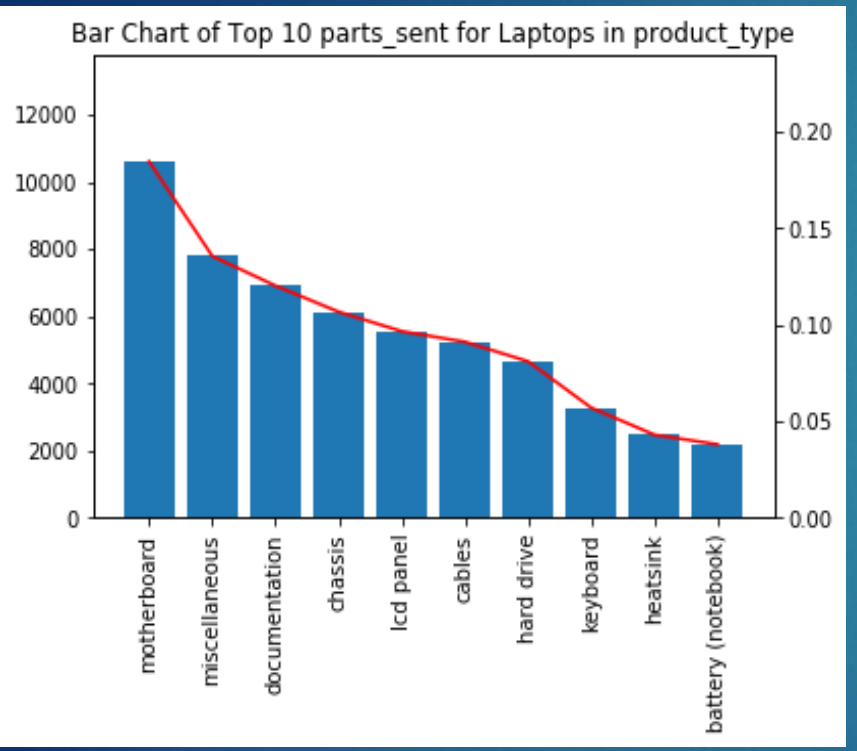
# The Issues needing to Contact Managers

## Topics Identified by Agents

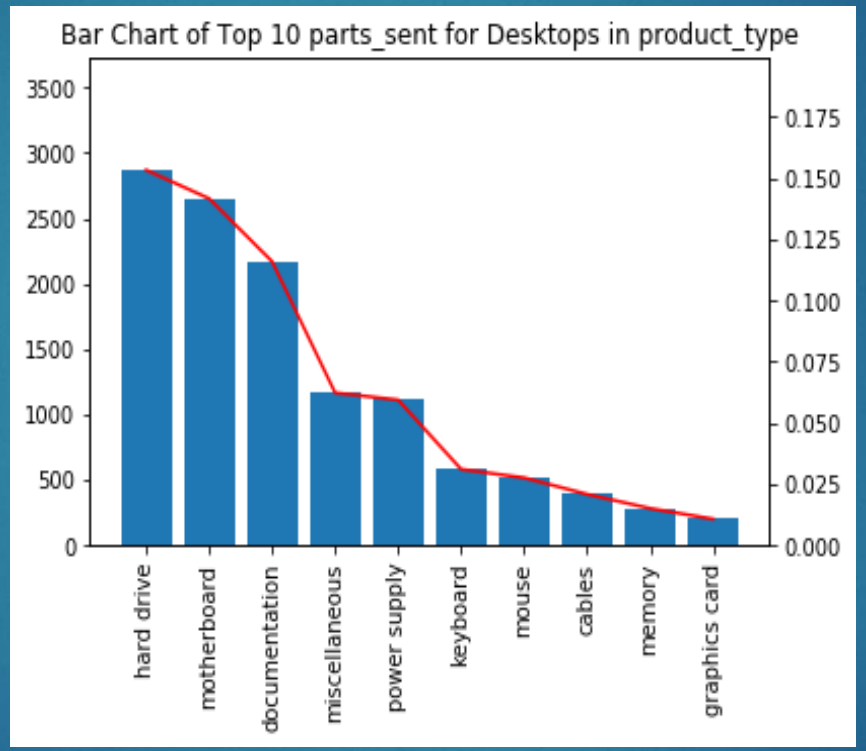


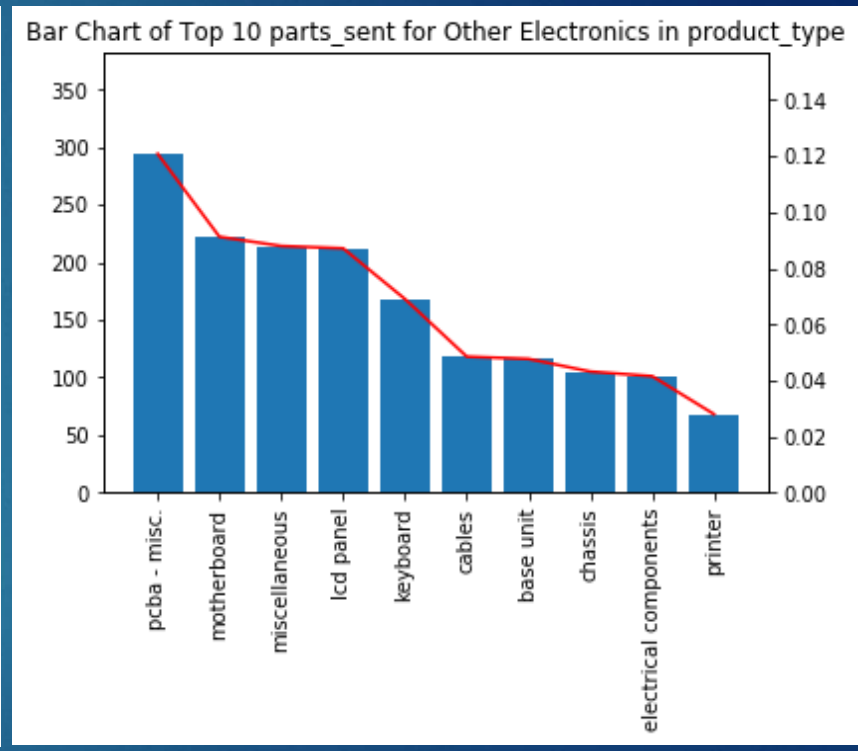
# Top Sent Parts for Different Product Type




Laptop




Desktop

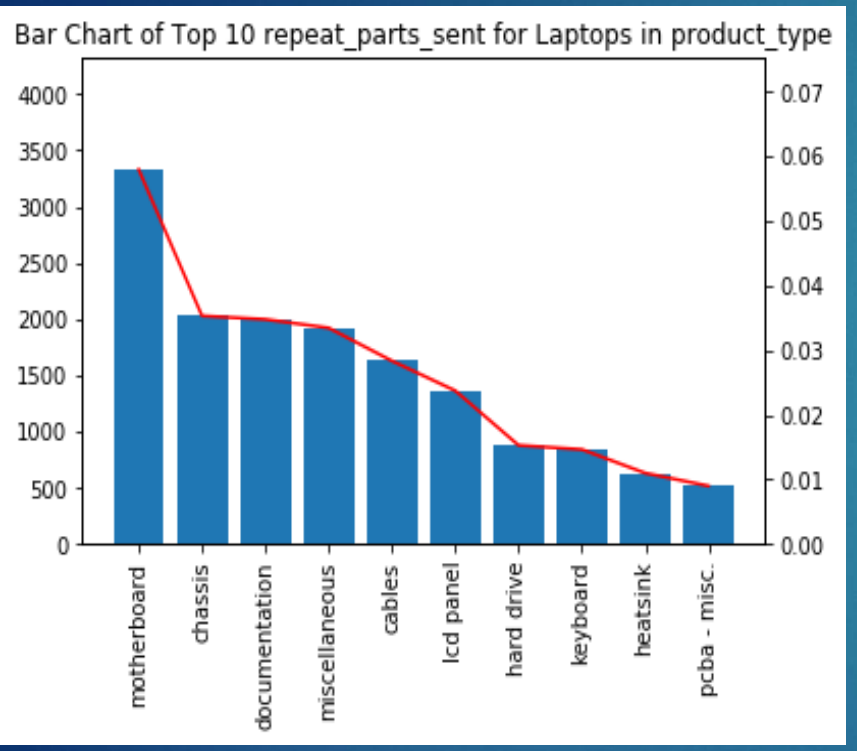



Other Electronics

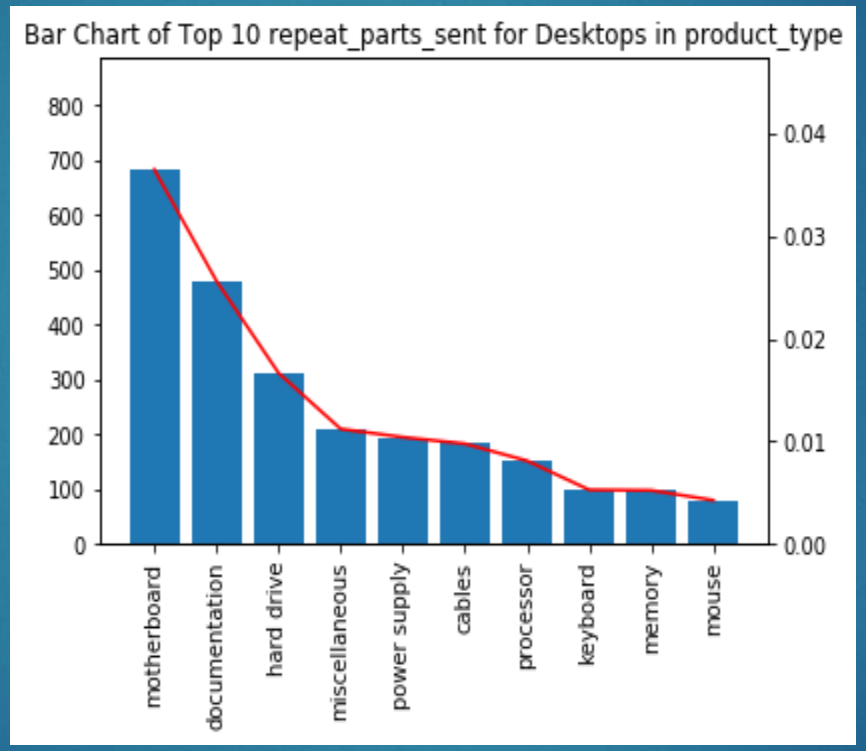
*Motherboard was the main problem requiring more part sent, and documentations were always something customers required

** As observed in the topics analysis, hard drive is another big problem for desktop

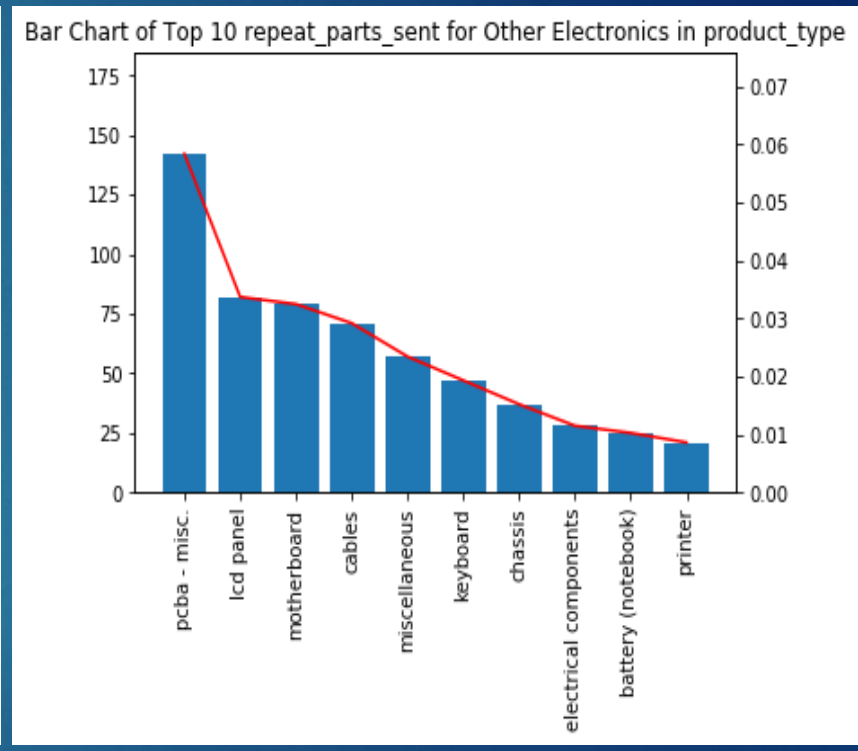
# Top Re-sent Parts for Different Product Type




Laptop




Desktop

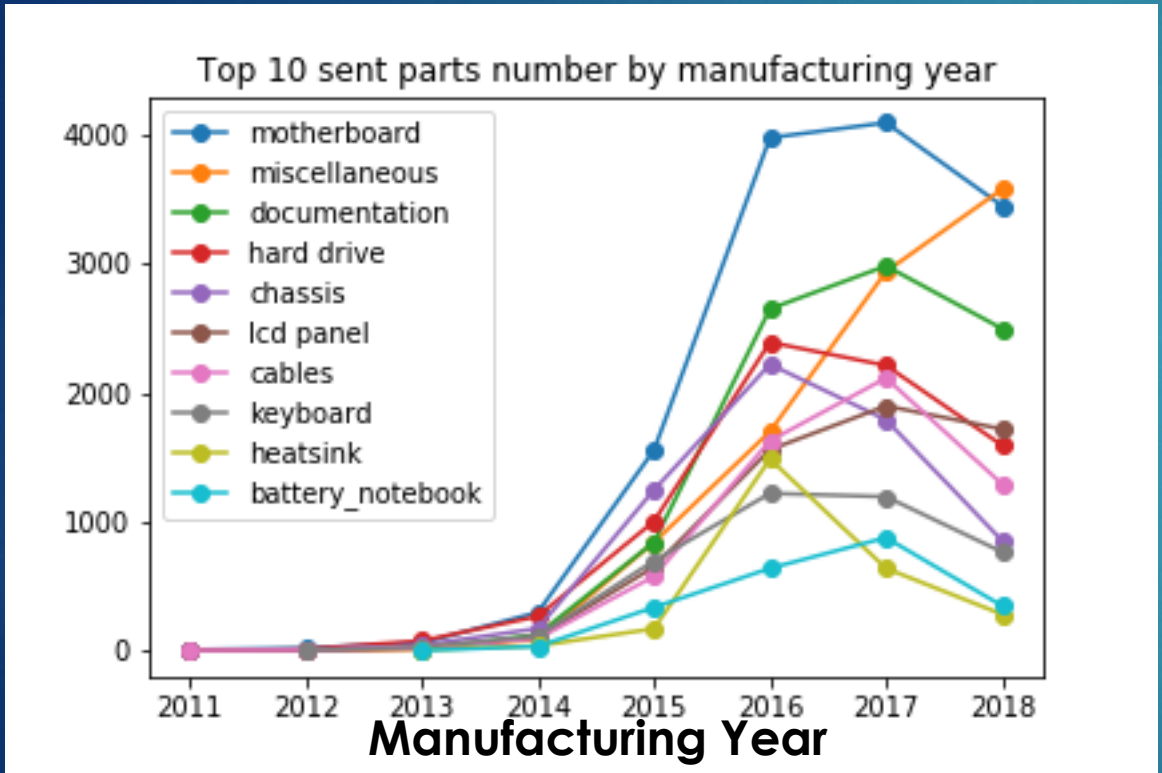



Other Electronics

*Motherboards is still the top of the list

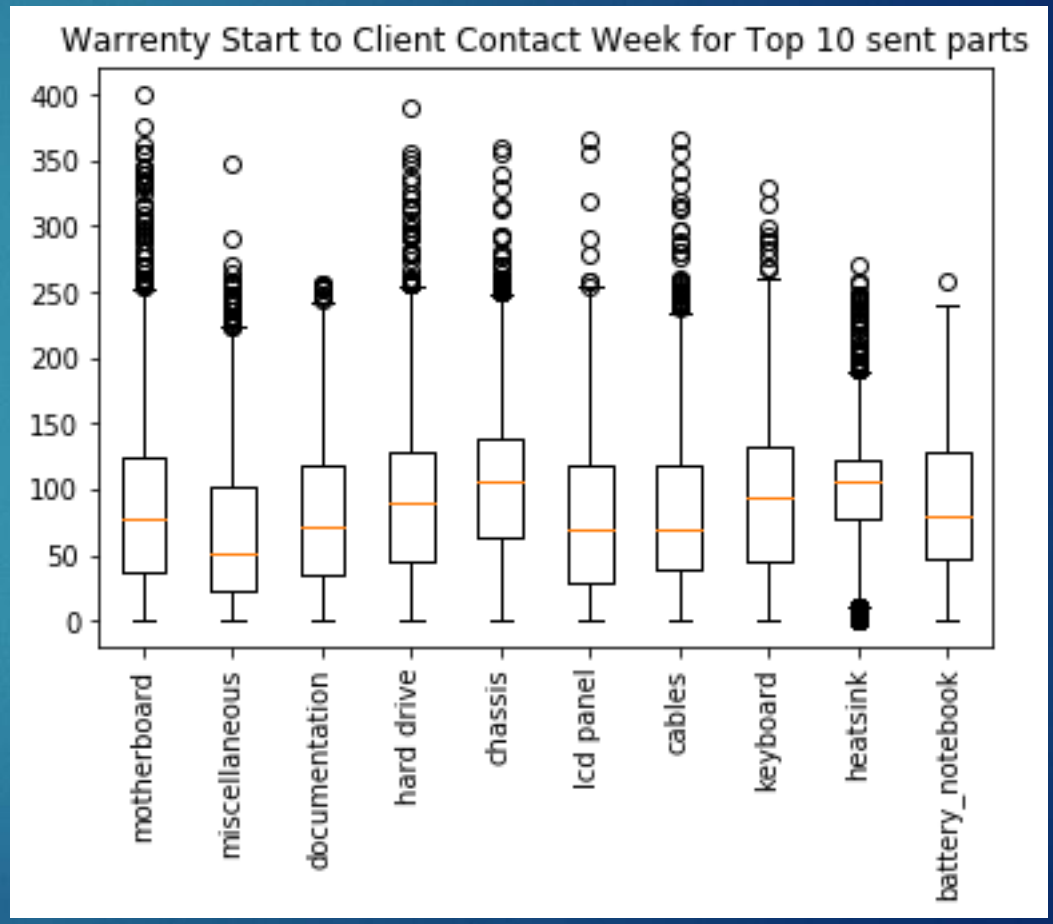
# Top Sent Parts by Year and Duration

## Number of top 10 sent part by year




*Same here needs to sales number to get the proportion
** Motherboard is always dominating the list while need to pay attention to hard drive, LCD and documentations

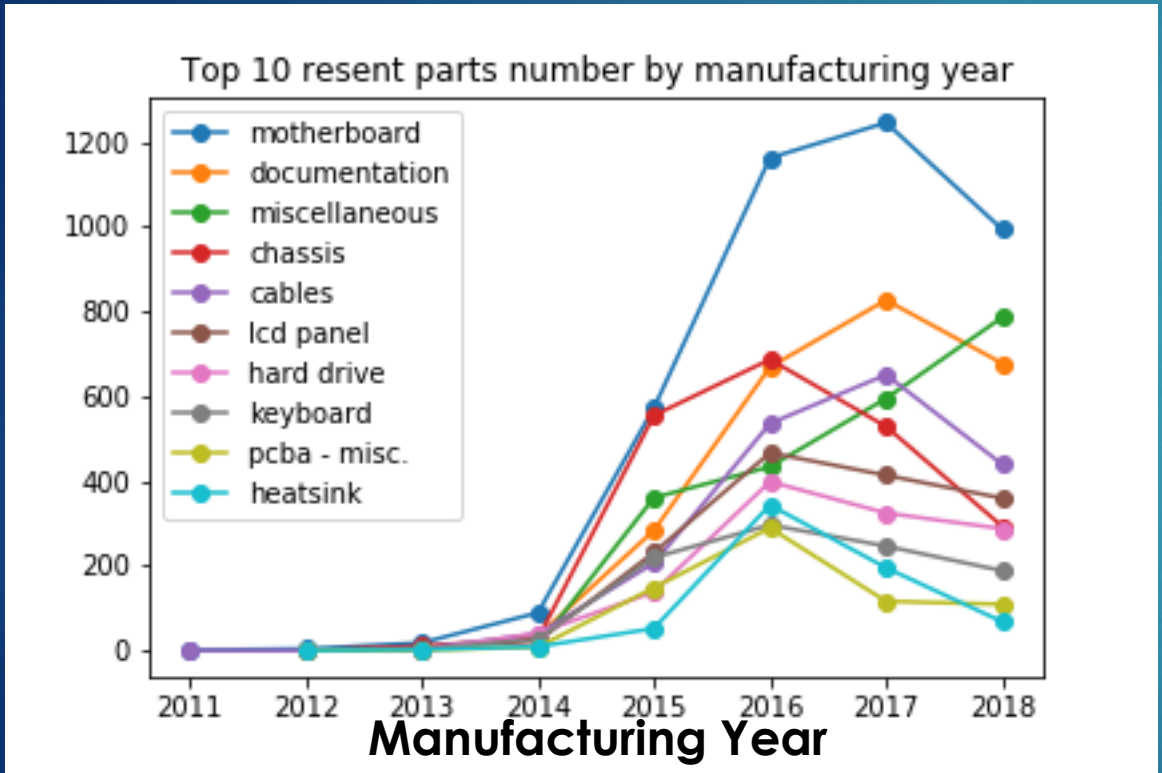
## How soon customer contacted after the warranty started



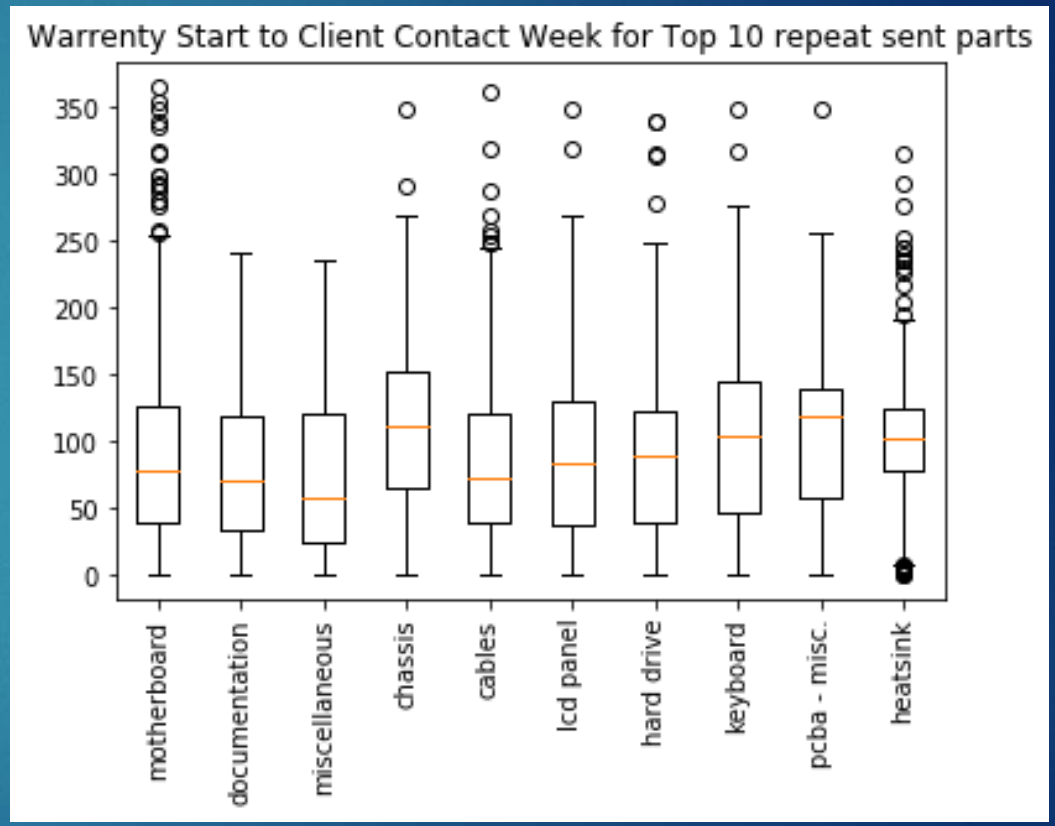

Need to pay attention to LCD panel for having issues relatively sooner and the larger variation

# Top Re-sent Parts by Year and Duration

## Number of top 10 re-sent part by year

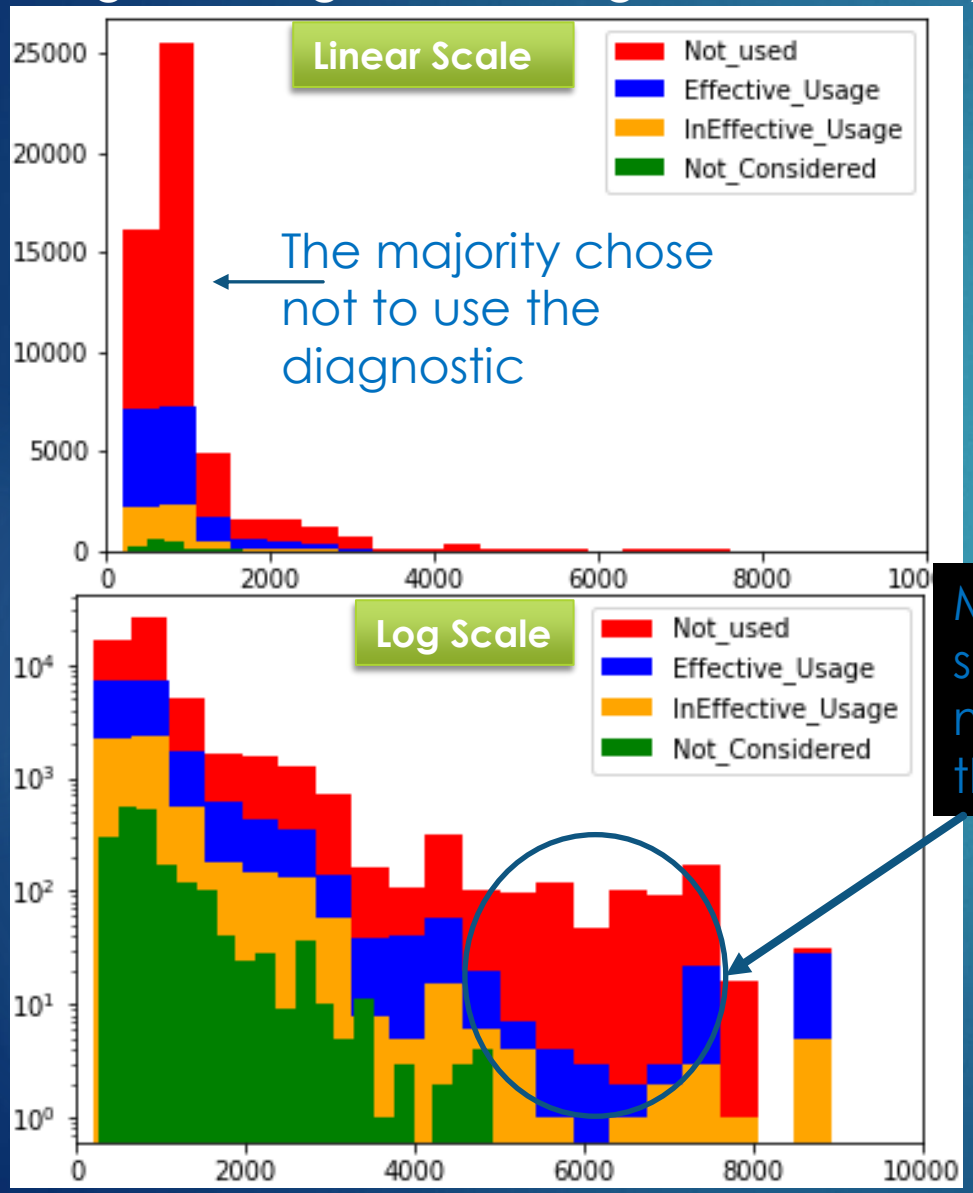



## How soon customer contacted after the warranty started



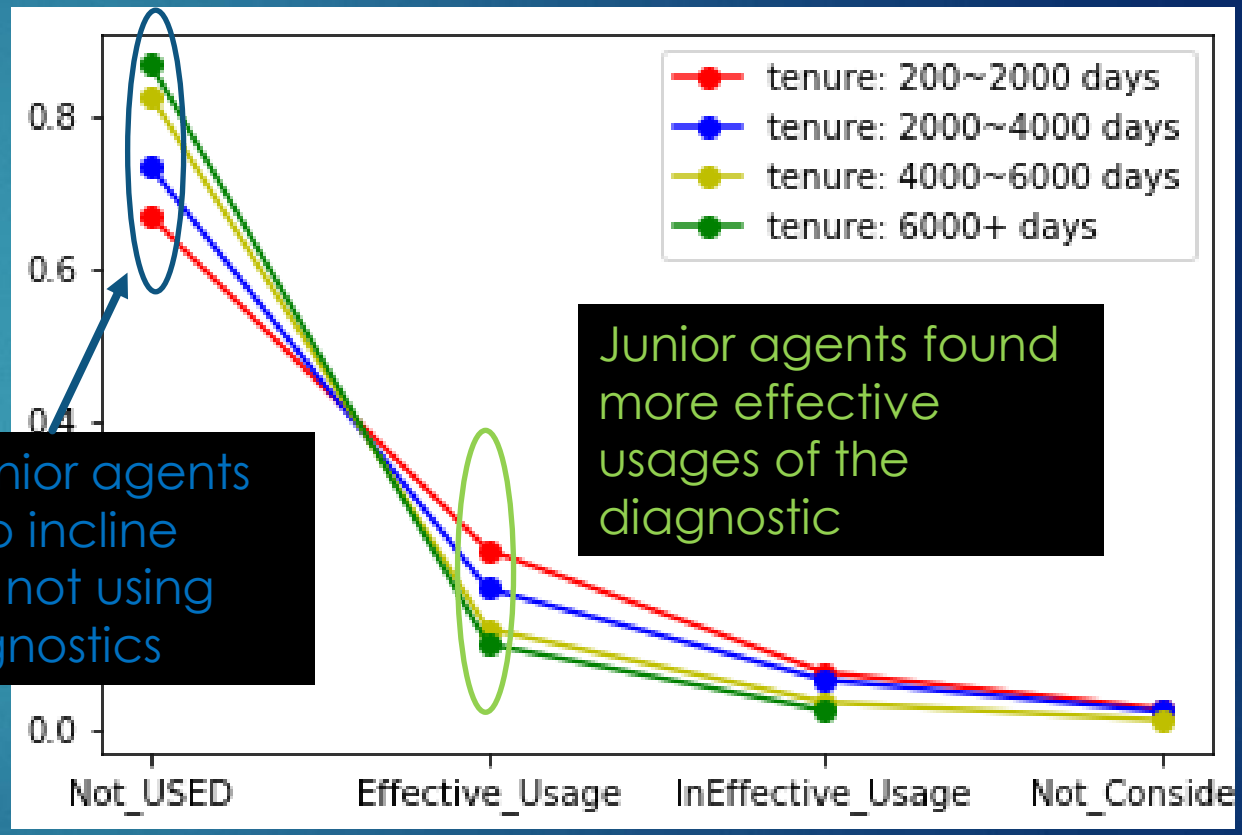
# Agent Tenue Day vs. Diagnostics

Usage of Diagnostics vs Agent Tenure Days




More senior agents seems to incline more to not using the diagnostics

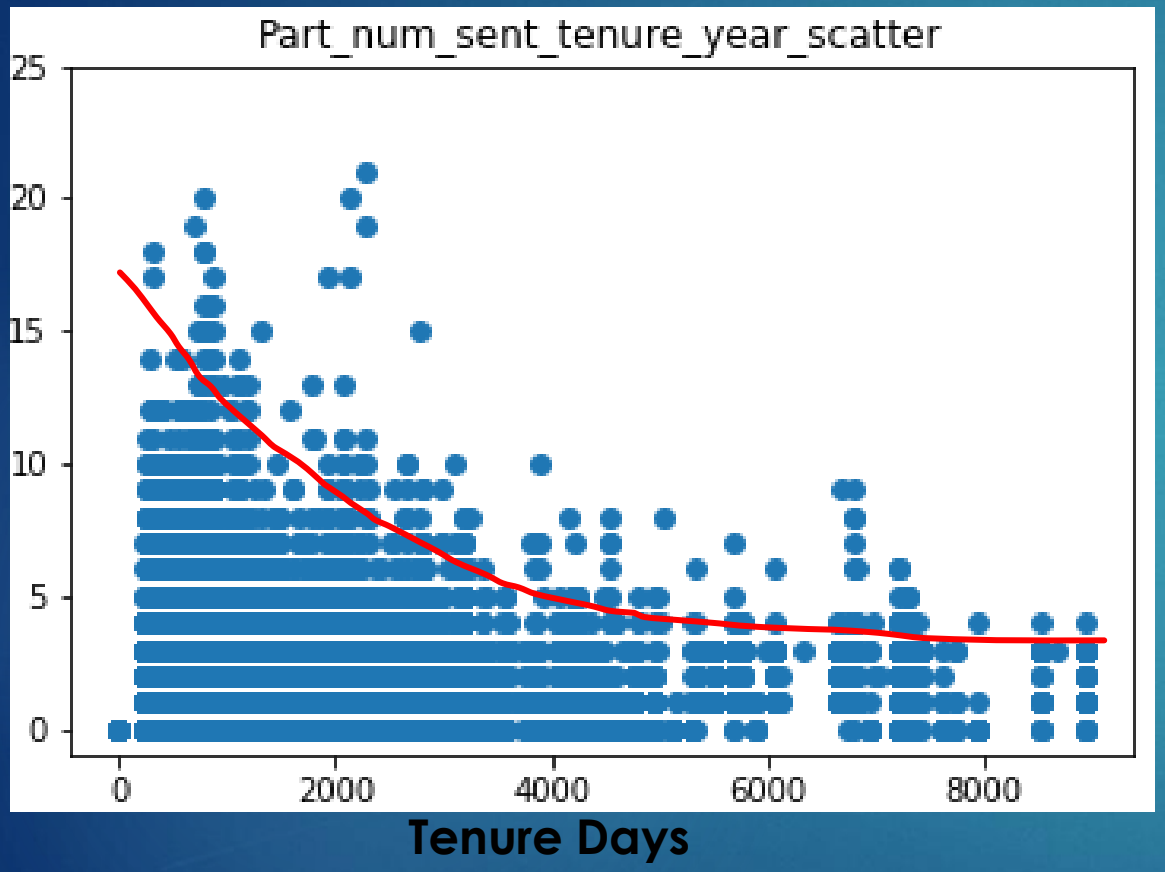
Proportion of Diagnostics vs Tenure Days Group (2000 days per group)



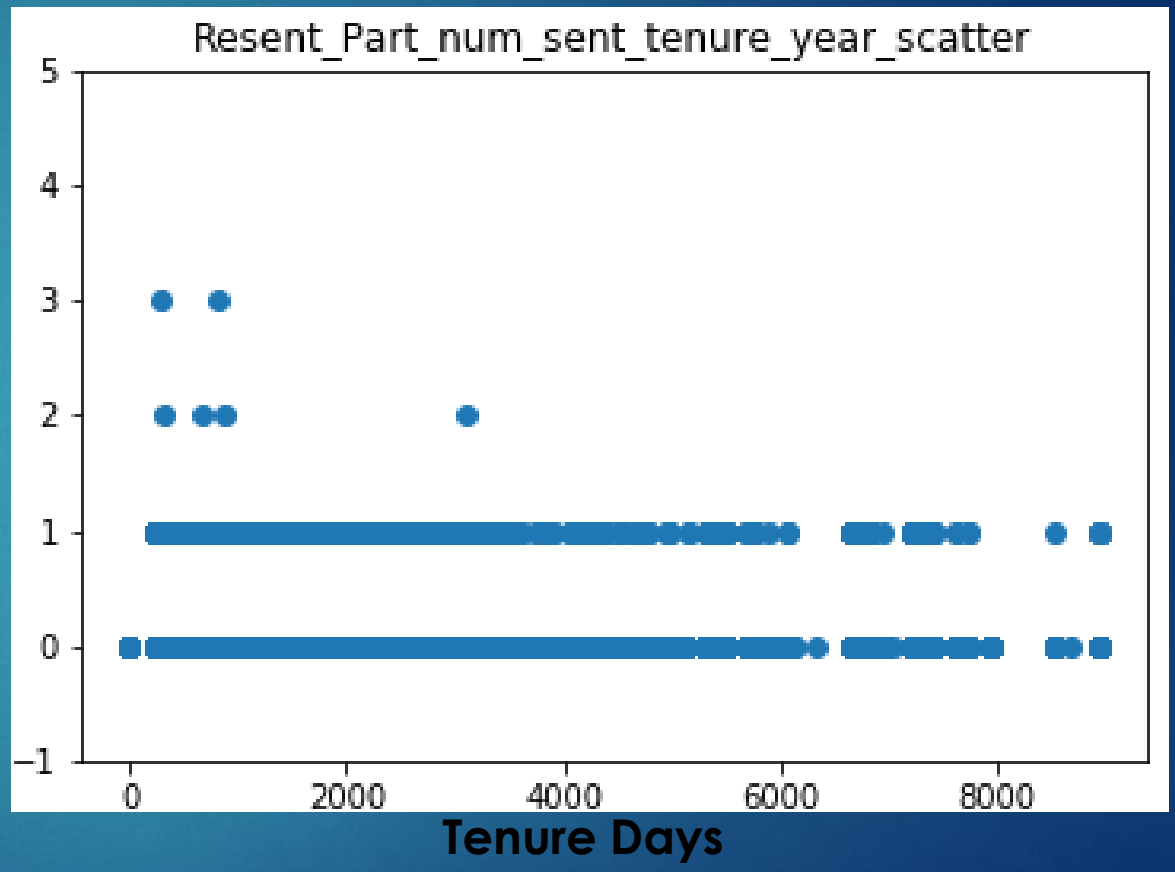

Junior agents found more effective usages of the diagnostic

# Part Number Sent vs Agent Tenure Days

Sent Part vs Tenure Days

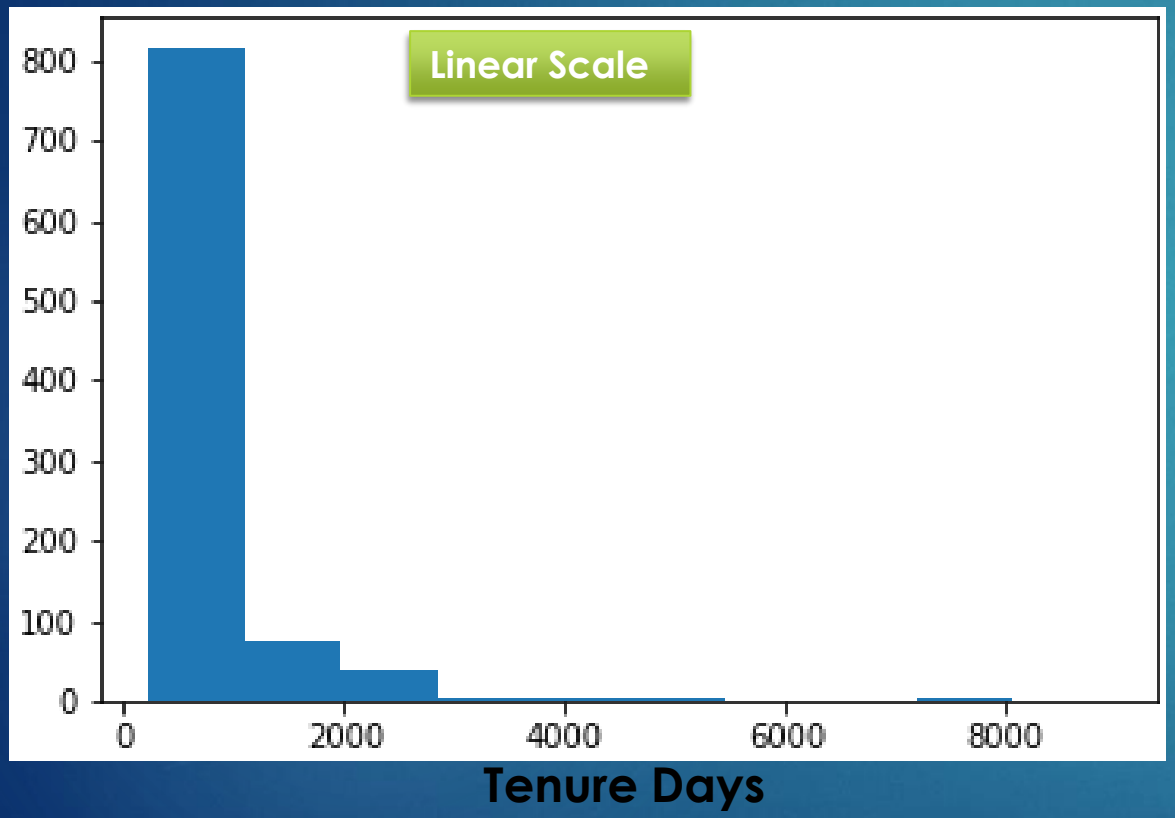


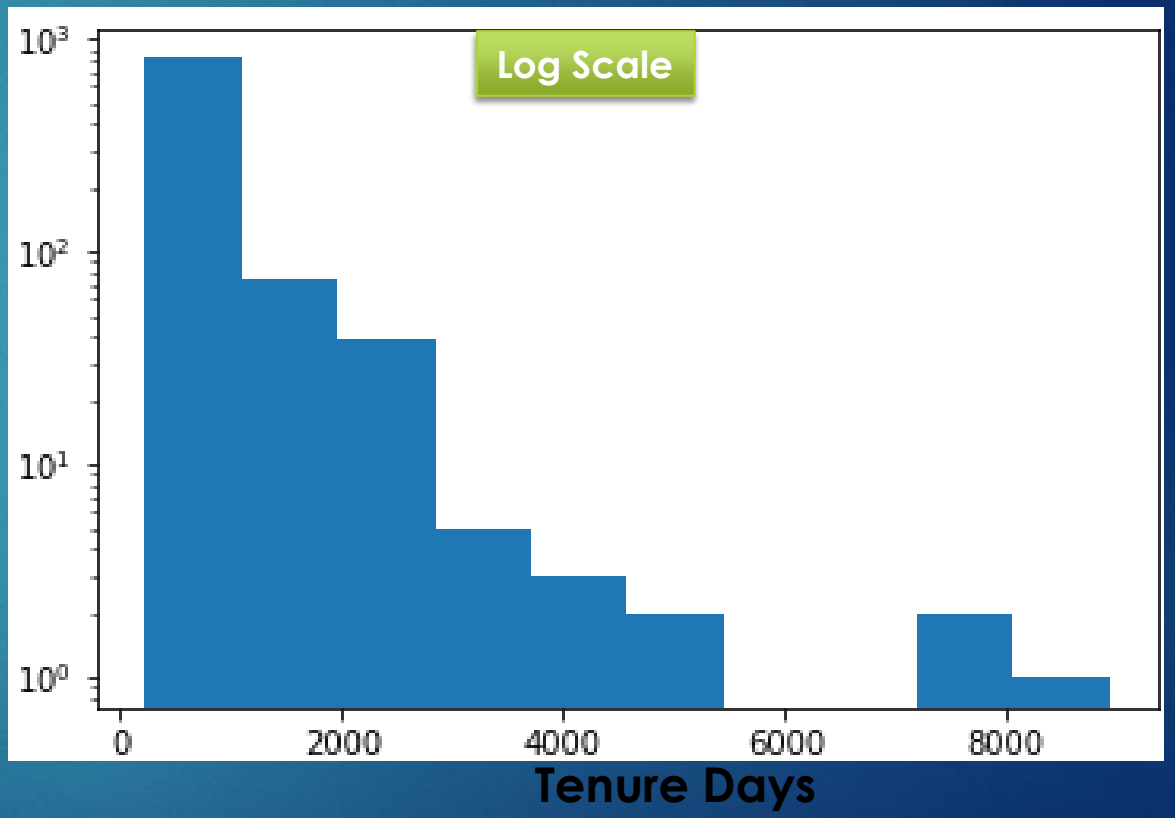

Re-Sent Part vs Tenure Days




***Junior agents tended to send more parts than senior agents**

# Tenure Day Distribution for Agents Contacting Manager




* 942 instances contacting managers

**Junior agents tended to contact manager more

# Summary and Observation

- The input data (all from week 40 of 2018) were cleaned by duplication removal, missing value QC and in accurate data removal
- For data exploration and analysis:
  - The main issues for customer contacts were booting/motherboard problem followed by hard drive, LCD and documentation resent
  - Some subsets having specific issues: ie. Desktop had more hard drive issue; internet/wireless issues standing out more at area of Milky Way
  - Around 50~100 weeks after warranty started were the time customer started to contact for issue. But need to check individual item/part with different variations specifically
  - Junior agents inclined more to use diagnostics and contact manager for help while they also tended to consume more parts when fixing issues (more training needed for the junior agents and need to encourage senior agents to use the diagnostics more)
- Some questions:
  - For the time/year analysis, it would be helpful to have the year sale for different items/product types to extract the proportion information. So that we can QC the quality changes of different products

# Overview

- Here is the general layout of this presentation
- Analysis on Products and Customer Services:
  - Initial Data QC and Data Cleansing
  - Data Exploration and Analysis (focusing on products and services)
  - Observation, Conclusion and More Questions
- **Exploring Graph DB with the Same Data Set**

# Exploring Graph DB

- Explored the Graph DB with Dell customer services data
- Software: Neo4j Desktop (1.3.4) and the DBMS was built on local machine (originally considered using AuraDB. But latter switch back to local machine since the data set is not big. Moreover, AuraDB has node and relation number limit)
- py2neo was another option but haven't chance for more exploration.
- Used the cleaned data set (instead of the raw) as input for this work.

# Import the Data and Building Nodes/Relationships

- The whole process including importing data and creating nodes/relationships was finished in one round. Here it was broken into setps.
- Import data:
  - Using periodic input with every 1000 rows
  - Seems quite slow with the LOAD CSV comments (it took several mins to load such a small data and almost 30 mins to build the all the nodes and relationships
  - Cypher may be faster

```
//Using Periodic Input with every 1000 rows
:auto USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS FROM 'file:///data_bk3_af_drop_time.csv' as line
with line
```

# Import the Data and Building Nodes/Relationships

- Building nodes in 2 stages:
  - Build basic nodes: instance/record, product, agent

```
//create instance/record nodes with product ID, issue and contact type as properties
MERGE (inst:instance {id: line.asst_id,
issue: line.issue_type,
contact_type: line.contact_type})

//create product nodes with different properties.
//Some of the properties will be later reloaded as node for segmentation relation QC
MERGE (pd:product
{pd_id: line.asst_id,
type: line.product_type,
maunf_wk: line.mnfture_wk,
contract_st: line.contract_st,
contact_wk: line.contact_wk,
contact_contract_wk: line.contact_contract_wk,
topic: line.topic_category})

//create agent nodes with tenure days. Also use tenure days as ID
MERGE (ag:agent {
    ag_id: line.agent_tenure_indays,
    tenure: line.agent_tenure_indays})
```

# Import the Data and Building Nodes/Relationships

- Building nodes in 2 stages:
  - Build basic nodes: instance/record, product, agent
  - Build categorical nodes for more relationship display

```
27 //create nodes with the different categorical data
28 MERGE(pb:problem {pb_type: line.topic_category}) //Topics node
29 MERGE (pdt:product_type {Name: line.product_type}) //Product type node
30 MERGE (rg:region {Name: line.region}) //region node
31 MERGE (cntr:country {Name:line.country}) // country node
32 MERGE (mnfyr:mnf_yr {Year: line.mn_year}) //manufacturing year node
33 MERGE (contryr:contract_st_yr {Year: line.contract_st_year}) //contract start year node
34 MERGE (conttype:contact_type {Name: line.contact_type}) // customer contact type node
```

# Import the Data and Building Nodes/Relationships

- Building Basic Relationship
  - Instance ‘report’ product
  - Agent ‘report’ product

```
36 //create relationship: instance/record 'reporting about' product
37 MERGE (inst)-[:report]→(pd)
38
39 //create relationship: agents 'fix' product.
40 //Include some part, diagnostics and manager flag properties
41 MERGE (ag)-[fx:fix]→(pd)
42 SET fx.fix_type = line.repair_type,
43 fx.diagnostics = line.diagnostics,
44 fx.part_sent = SPLIT(line.parts_sent,','),
45 fx.part_count = line.parts_ct,
46 fx.rp_part_sent = SPLIT(line.repeat_parts_sent,','),
47 fx.rp_part_count =line.repeat_ct,
48 fx.manage_flg = line.contact_manager_flg
```

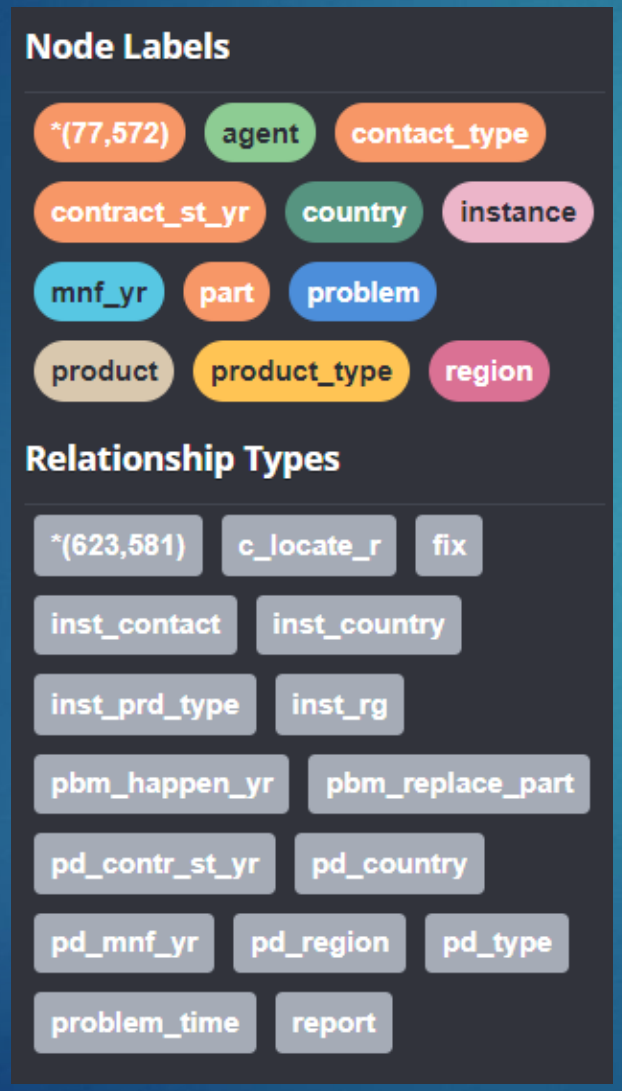
# Import the Data and Building Nodes/Relationships

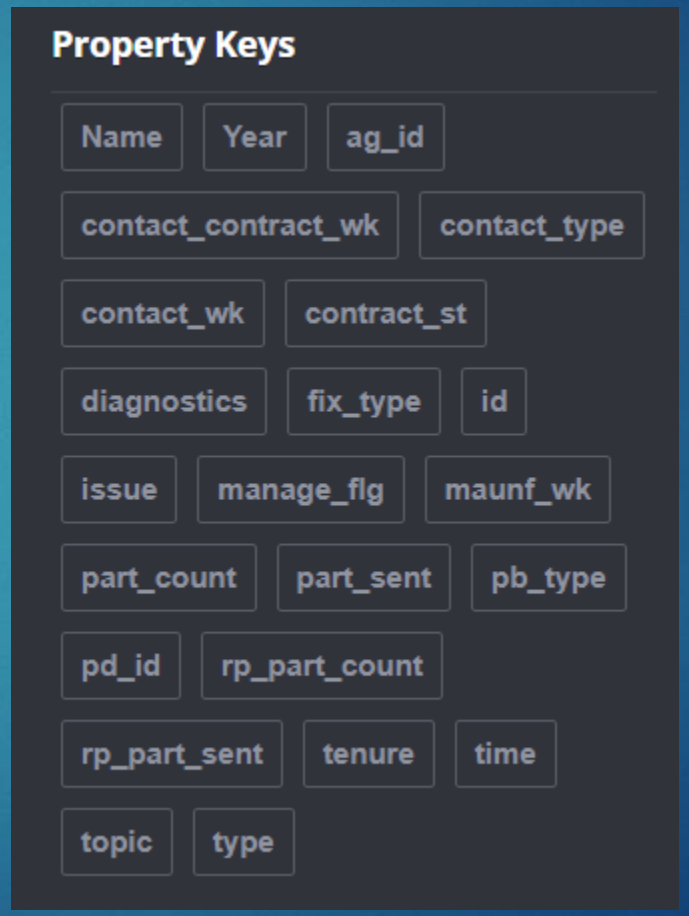
- Building Other Relationship
  - Products -> categorical nodes
  - Instances -> categorical nodes

```
50 //create relationship for product with different categorical data
51 MERGE (pd) -[:problem_time {time: line.contact_contract_wk}]→(pb)
52 merge (pd) -[:pd_type]→(pdt)
53 merge (pd) -[:pd_region]→(rg)
54 merge (pd) -[:pd_country]→(cntr)
55 merge (pd) -[:pd_mnf_yr]→(mnfyr)
56 merge (pd) -[:pd_contr_st_yr]→(contryr)
57
58 //create relationship for instances/records with different categorical data
59 merge(inst) -[:inst_prd_type]→(pdt)
60 merge(inst) -[:inst_rg]→(rg)
61 merge(inst) -[:inst_country]→(cntr)
62 merge(inst) -[:inst_contact]→(conttype)
63
64 //create relationship: country 'locate at' region
65 MERGE (cntr)-[:c_locate_r]→(rg)
```

Extra country ‘locate’ region node

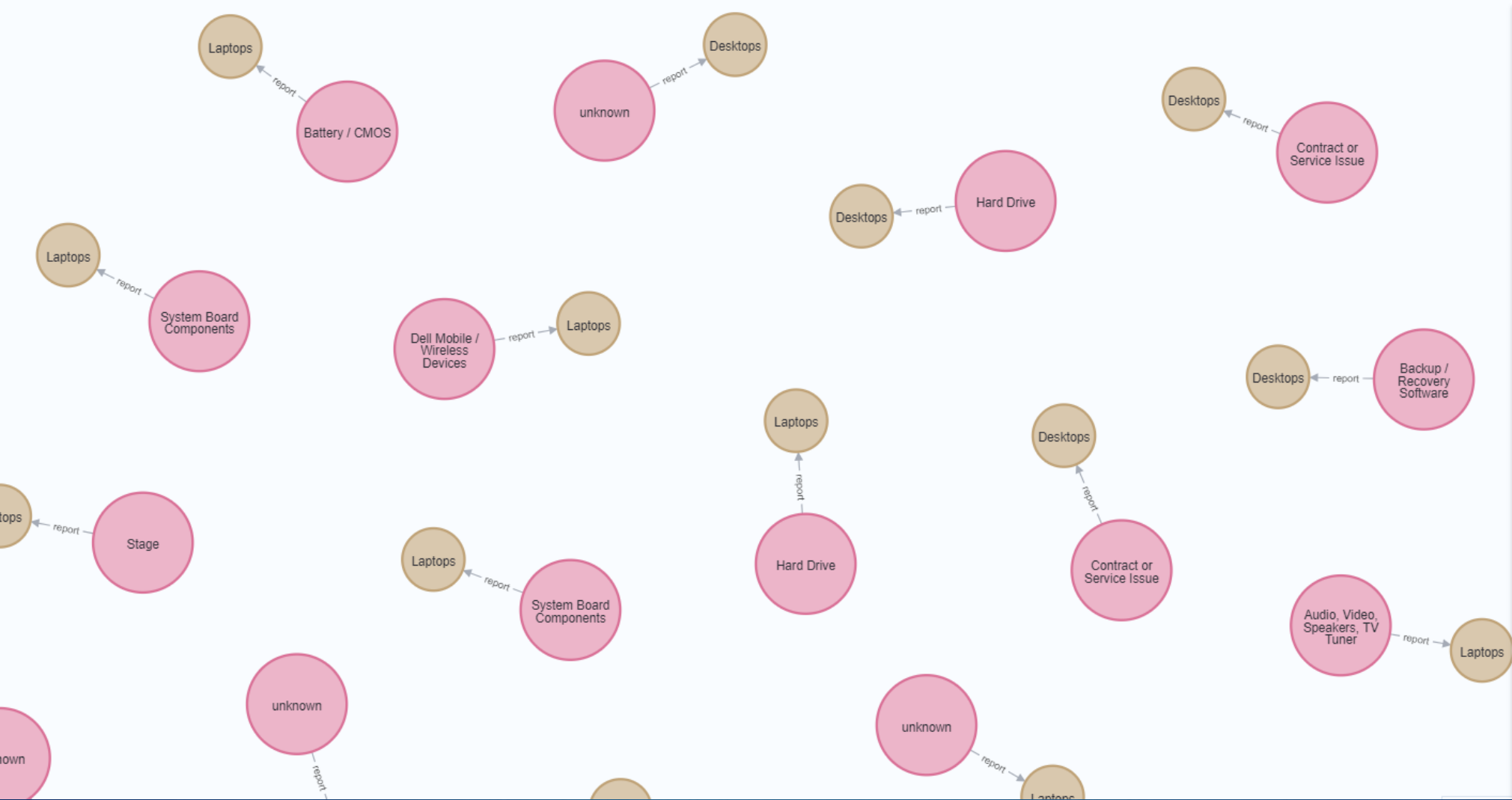
# Model Build: Node/Relationship/Properties



# Graph Display Example: Instances 'report' Products

```
MATCH p=()-[r:report]→() RETURN p LIMIT 25
```




Overview

**Node labels**

* (50) instance (25) product (25)

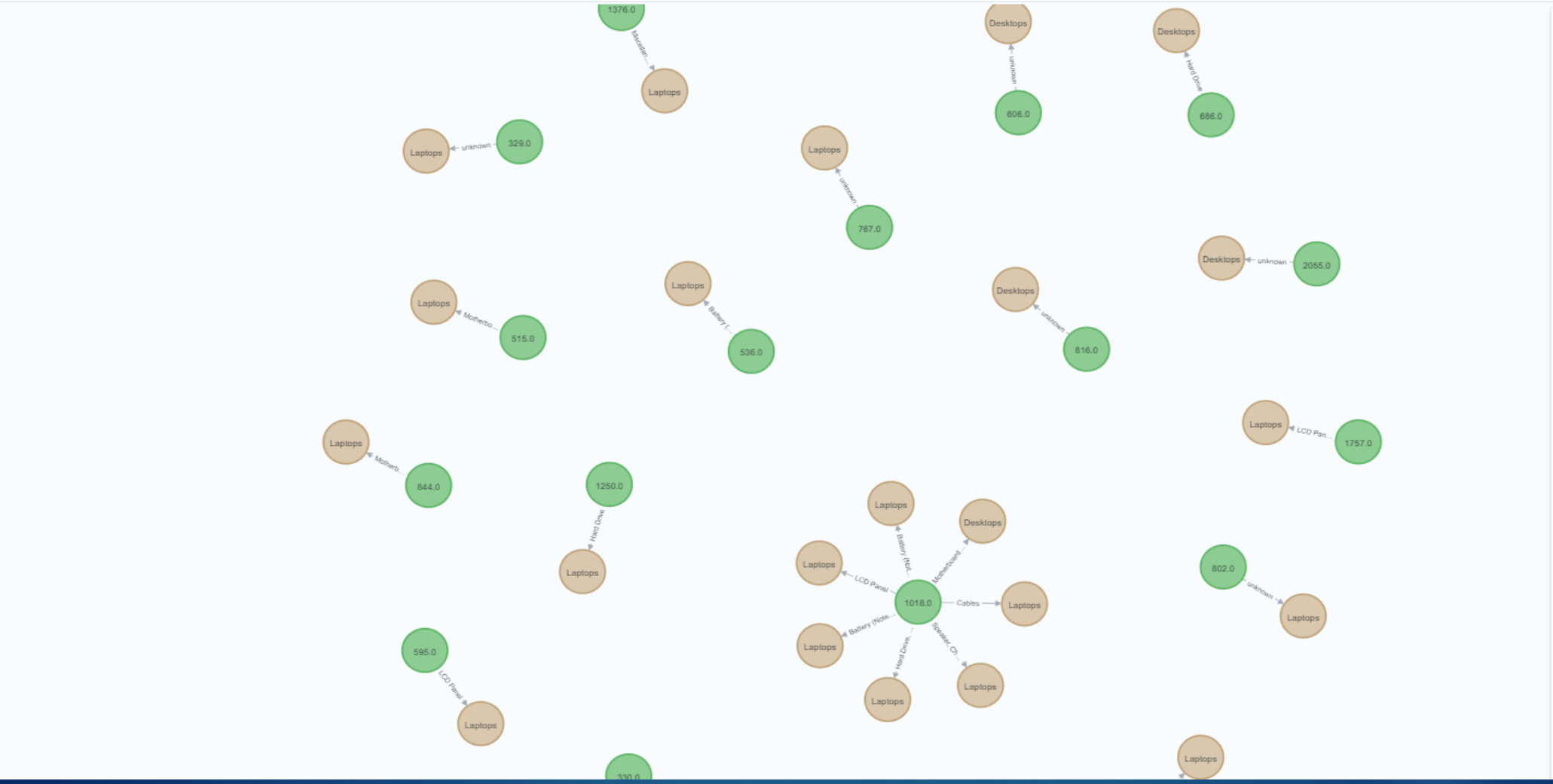
**Relationship Types**

* (25) report (25)

Displaying 50 nodes, 25 relationships.

# Graph Display Example: Agents 'fix' Products

```
MATCH p=()-[r:fix]→() RETURN p LIMIT 25
```




Overview

**Node labels**

* (44) agent (19) product (25)

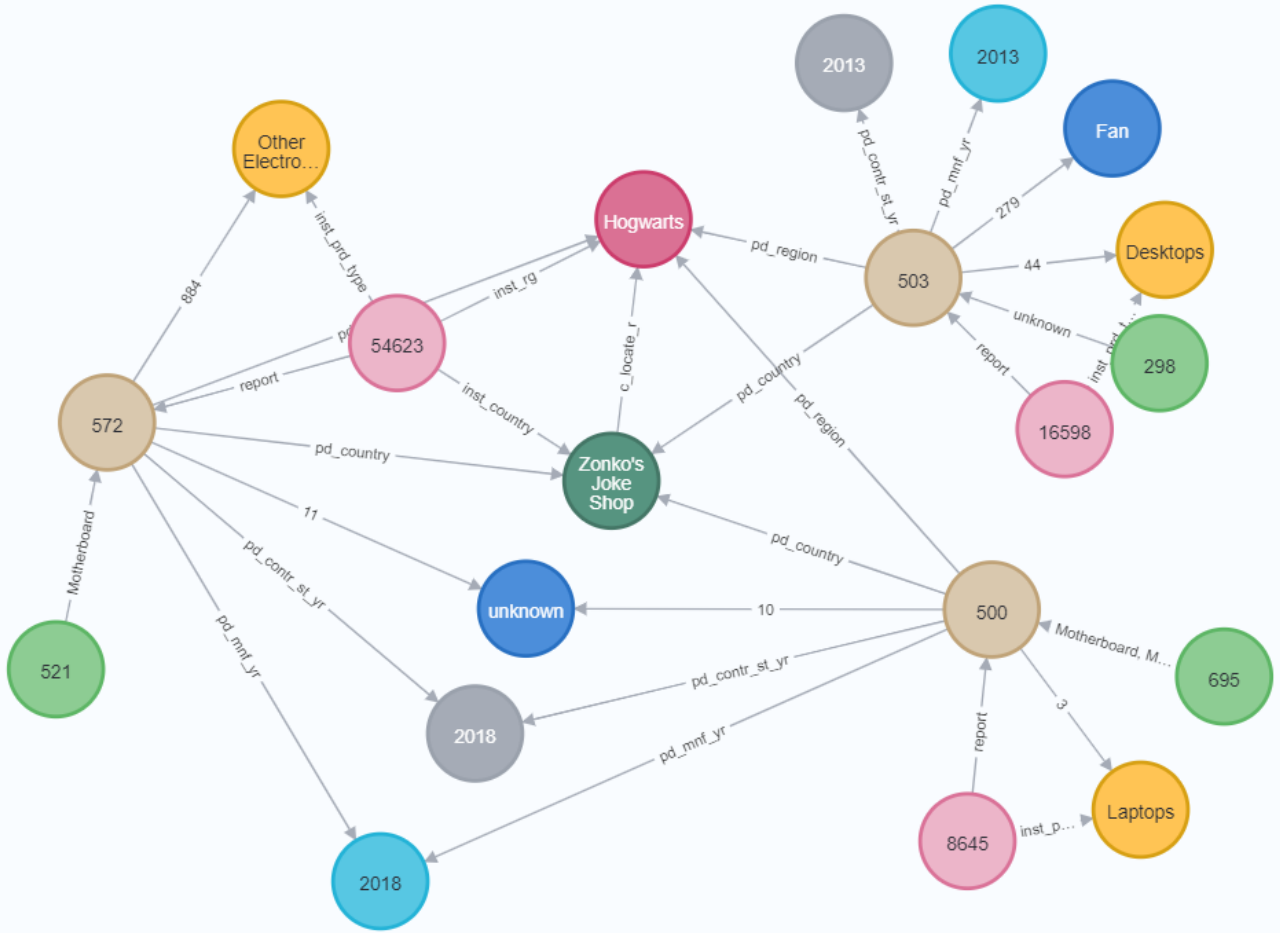
**Relationship Types**

* (25) fix (25)

Displaying 44 nodes, 25 relationships.

# Graph Display Example: Exploring Multiple Relationship with 3 Products

```
match (n:product {type:'Laptops'}),(m:product {type:'Desktops'}),(l:product {type:'Other Electronics'}) return n,m,l limit 1
```




Overview

**Node labels**

* (20) | product (3) | instance (3) | problem (2) | product_type (3) | region (1) | country (1) | mnf_yr (2) | contract_st_yr (2) | agent (3)

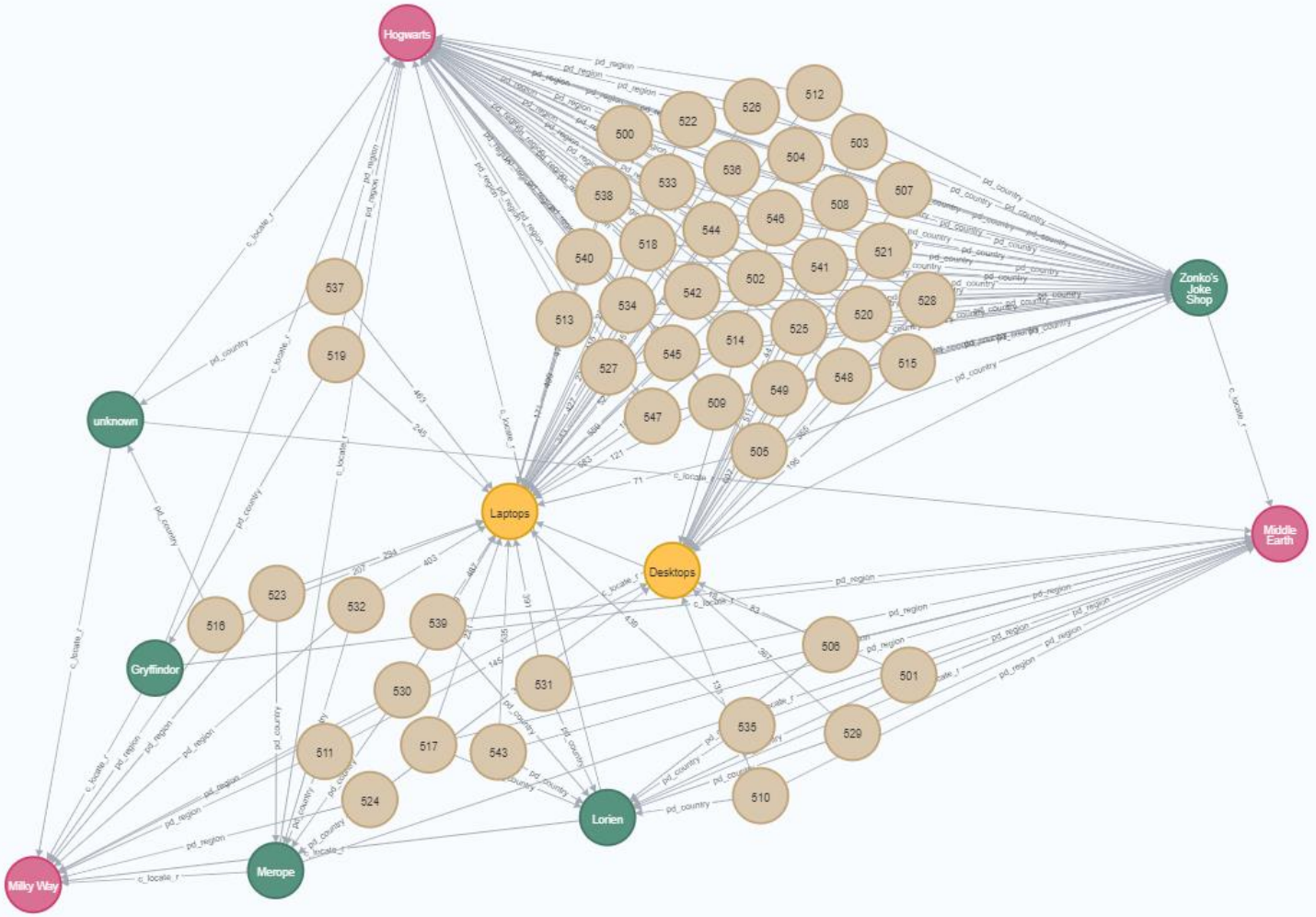
**Relationship Types**

* (30) | report (3) | problem_time (3) | pd_type (3) | pd_region (3) | pd_country (3) | pd_mnf_yr (3) | pd_contr_st_yr (3) | fix (3) | inst_country (1) | inst_rg (1) | inst_prd_type (3) | c_locate_r (1)

Displaying 20 nodes, 30 relationships.

# Graph Display Example: Exploring Multiple Relationship Focusing on 2 Main Product Types

```
match p=(pd:product) -[:pd_type | :pd_country | :c_locate_r]→() return p limit 100
```




Overview

**Node labels**

* (60) | product (50) | country (5) | product_type (2) | region (3)

**Relationship Types**

* (165) | pd_country (50) | pd_type (50) | c_locate_r (15) | pd_region (50)

Displaying 60 nodes, 165 relationships.

# Observations

- GRAPH DB is very power tool for visualization data relationship
- No joint is needed since the connection has been built in the data base
- Easy to use: the language is very similar to SQL. More exploration on Cypher and Py2neo to find faster and more versatile application with GRAPH DB
- The display is giving immediate information about nodes and relationships. But sometime may needs smarter and more concise display when the number of nodes involved is large the relation ship is complicated