# UNDERSTANDING CUSTOMERS BETTER THROUGH NEURAL NETWORK EMBEDDINGS

Adam Hornsby
(adam.hornsby@dunnhumby.com)



dunnhumby

# ABOUT ME

# ABOUT ME

- Senior Data Scientist @ dunnhumby (4.5 years)

# ABOUT ME

- Senior Data Scientist @ dunnhumby (4.5 years)
- **Researching and deploying** machine learning at scale

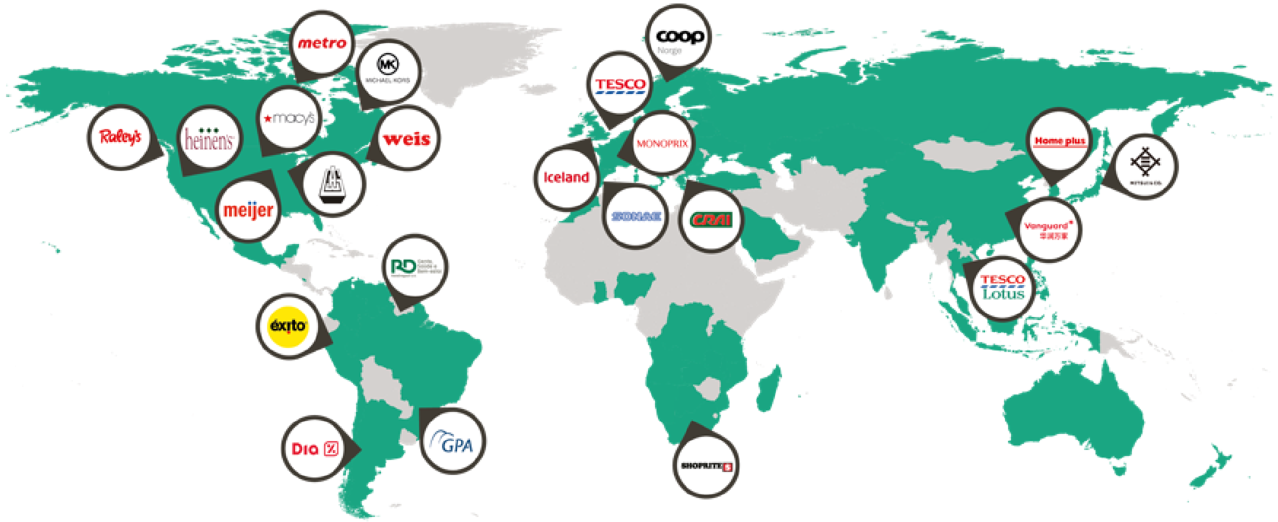
# ABOUT ME

- Senior Data Scientist @ dunnhumby (4.5 years)
- **Researching and deploying** machine learning at scale
- Experimental Psychology, PhD Student @ UCL (part-time)

# ABOUT ME

- Senior Data Scientist @ dunnhumby (4.5 years)
- **Researching and deploying** machine learning at scale
- Experimental Psychology, PhD Student @ UCL (part-time)
- Researching **consumer decision-making** through experiments, big-data and computational modelling

# EXTENSIVE HERITAGE AND EXPERIENCE WORKING WITH RETAILERS AND BRANDS




Leading Global FMCG Companies

L'ORÉAL, Coca-Cola, General Mills, Kellogg's, Unilever, Reckitt Benckiser, Mondelēz International, Nestlé, Henkel, PEPSICO

Leading Consumer Brands

ESTĒE LAUDER, MICHAEL KORS, LUXOTTICA, GUESS by Marciano, CHANEL, Ralph Lauren, Levi's, Kenneth Cole, L'ORÉAL, COACH New York

**Over 25** years experience

Using world-leading data and science to drive growth

**72** retailers and **1,000+** brands

# TODAY

# TODAY

1. Why **item similarity** is so important in retail science

# TODAY

1. Why **item similarity** is so important in retail science
2. What are **embeddings**?

# TODAY

1. Why **item similarity** is so important in retail science
2. What are **embeddings**?
3. How can **2vec** algorithms help?

# TODAY

1. Why **item similarity** is so important in retail science
2. What are **embeddings**?
3. How can **2vec** algorithms help?
4. Conclusions

# TODAY

1. Why **item similarity** is so important in retail science
2. What are **embeddings**?
3. How can **2vec** algorithms help?
4. Conclusions

(Massive thanks to Josh Cooper for doing lots of the thinking in this presentation)

# ITEM SIMILARITY

# HOW SIMILAR ARE THESE TWO PRODUCTS?


# HOW SIMILAR ARE THESE TWO PRODUCTS?



Questions of similarity are **everywhere in retail:**

# HOW SIMILAR ARE THESE TWO PRODUCTS?



Questions of similarity are **everywhere in retail:**

"Your selected product X **goes well with** product Y" (i.e. product complementarity)

# HOW SIMILAR ARE THESE TWO PRODUCTS?



Questions of similarity are **everywhere in retail:**

"Your selected product X **goes well with** product Y" (i.e. product complementarity)
"Your product X **was not available, so how about alternative** Y?" (i.e. product substitutability)

# HOW SIMILAR ARE THESE TWO PRODUCTS?



Questions of similarity are **everywhere in retail:**

"Your selected product X **goes well with** product Y" (i.e. product complementarity)
"Your product X **was not available, so how about alternative** Y?" (i.e. product substitutability)
"**Products of this type** tend to be placed together on the shelf" (i.e. product similarity)

# HOW SIMILAR ARE THESE TWO PRODUCTS?



Questions of similarity are **everywhere in retail:**

"Your selected product X **goes well with** product Y" (i.e. product complementarity)
"Your product X **was not available, so how about alternative** Y?" (i.e. product substitutability)
"**Products of this type** tend to be placed together on the shelf" (i.e. product similarity)
"**Customers like you** also tend to buy Y" (i.e. customer similarity)

# HOW SIMILAR ARE THESE TWO PRODUCTS?

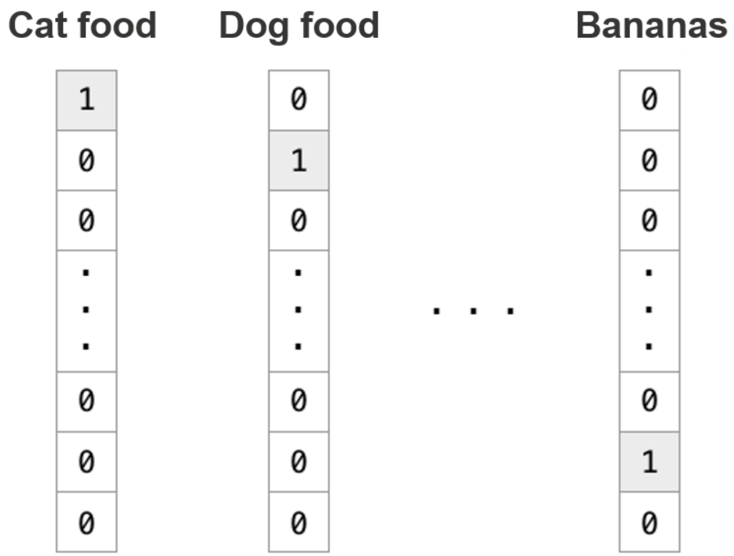


Questions of similarity are **everywhere in retail:**

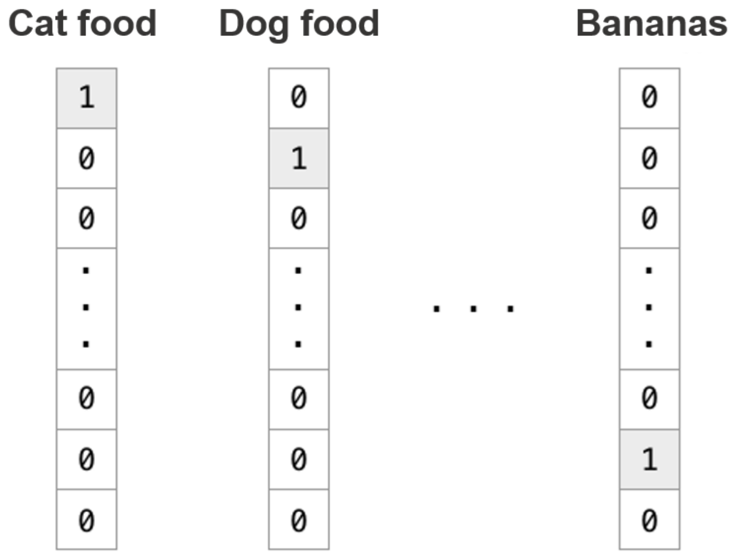
"Your selected product X **goes well with** product Y" (i.e. product complementarity)
"Your product X **was not available, so how about alternative** Y?" (i.e. product substitutability)
"**Products of this type** tend to be placed together on the shelf" (i.e. product similarity)
"**Customers like you** also tend to buy Y" (i.e. customer similarity)

Solving similarity is a **huge goal for data scientists** working to improve; recommendations, ranging, pricing, assortment and more

# TRADITIONAL METHODS DON'T HELP WITH SIMILARITY

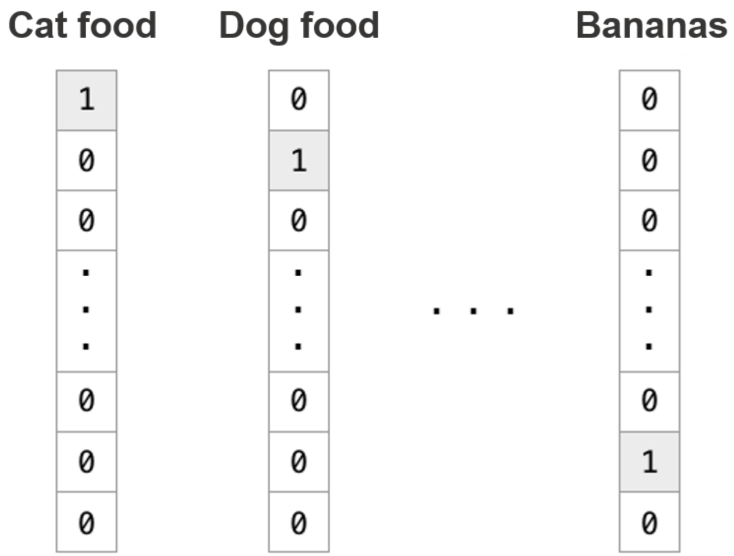


# TRADITIONAL METHODS DON'T HELP WITH SIMILARITY




Most scientists will use **dummy (or one-hot) encoding** to represent categorical data

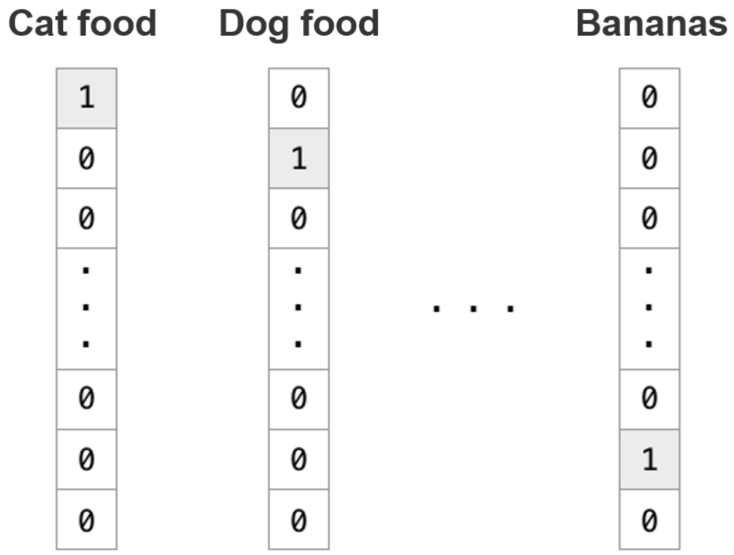
# TRADITIONAL METHODS DON'T HELP WITH SIMILARITY




Most scientists will use **dummy (or one-hot) encoding** to represent categorical data

Dummy coded data is **too sparse** (i.e. too many zeros)

# TRADITIONAL METHODS DON'T HELP WITH SIMILARITY



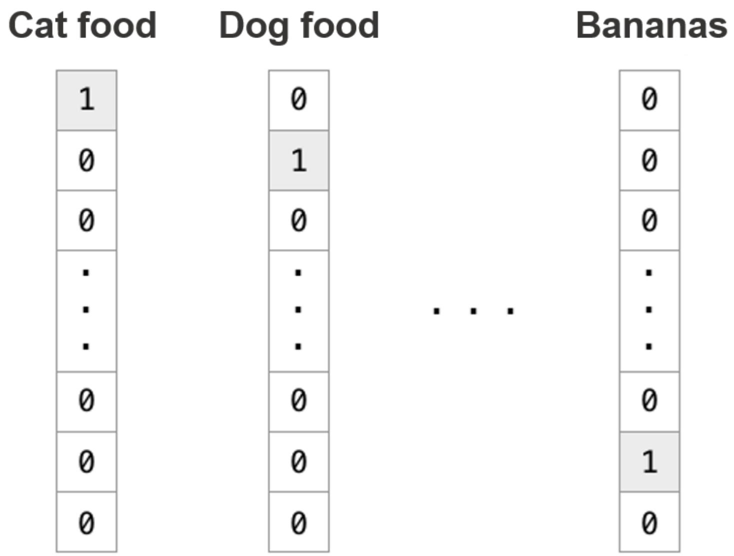

Most scientists will use **dummy (or one-hot) encoding** to represent categorical data

Dummy coded data is **too sparse** (i.e. too many zeros)
Any similarity measure will tell you that *cat food* and *dog food* are **totally unrelated** using this approach



dunnhumby

# TRADITIONAL METHODS DON'T HELP WITH SIMILARITY




Most scientists will use **dummy (or one-hot) encoding** to represent categorical data

Dummy coded data is **too sparse** (i.e. too many zeros)
Any similarity measure will tell you that *cat food* and *dog food* are **totally unrelated** using this approach

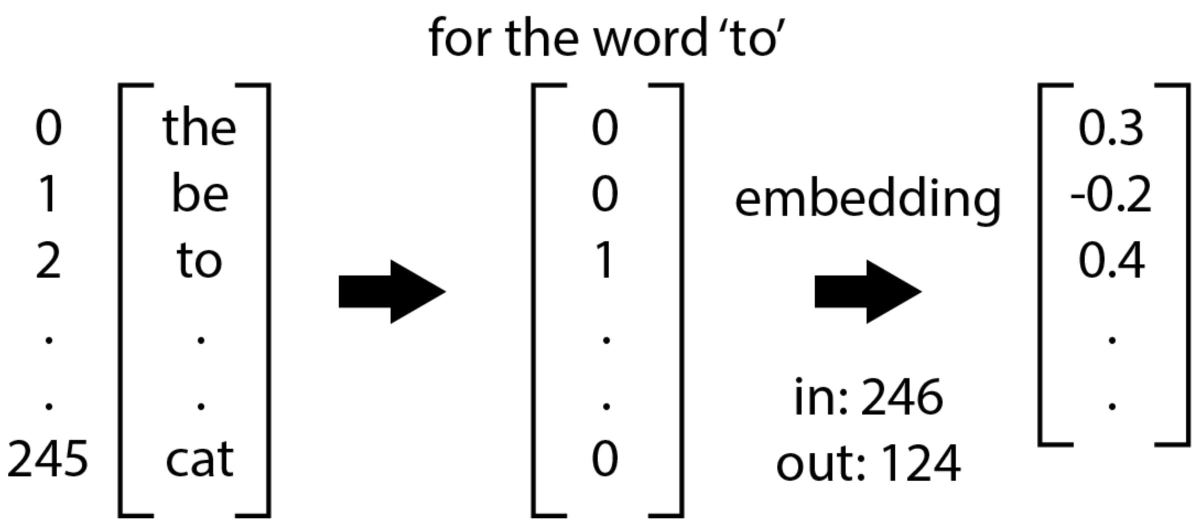
We need a way to represent these items using **dense** vectors



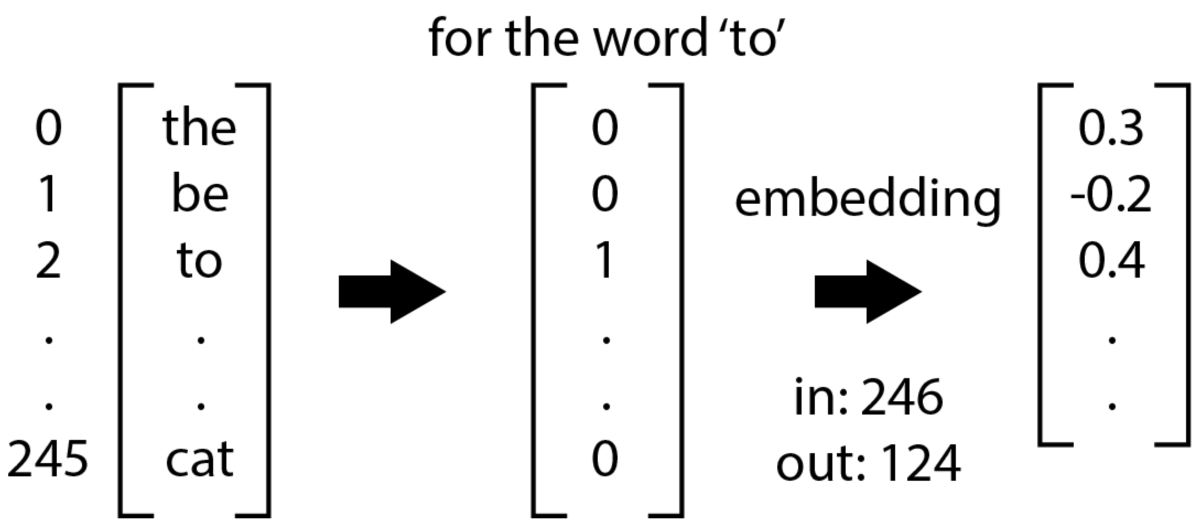
dunnhumby

# EMBEDDINGS

# EMBEDDINGS ARE DENSE REPRESENTATIONS

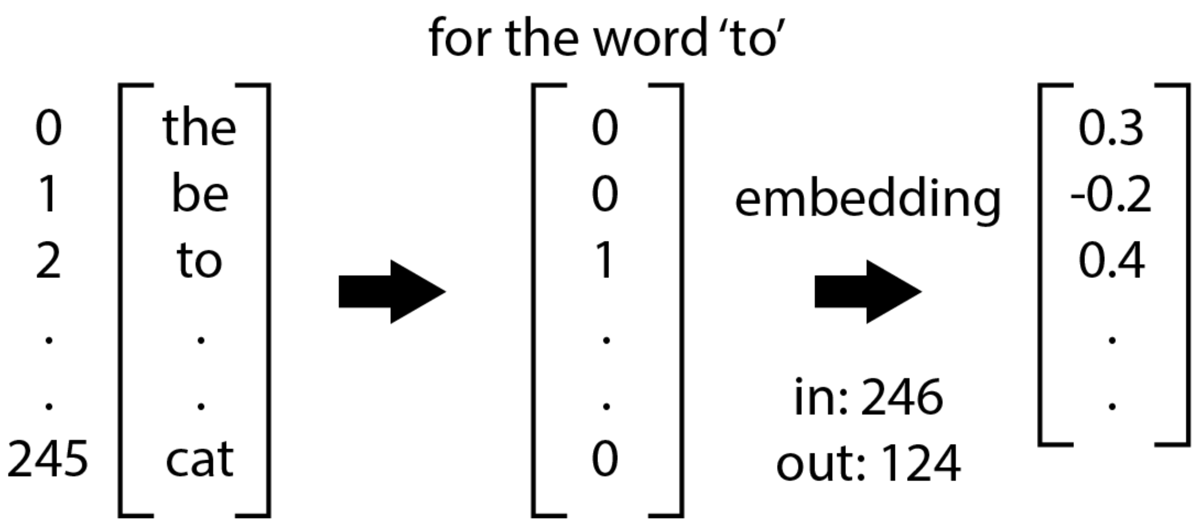


# EMBEDDINGS ARE DENSE REPRESENTATIONS




Traditionally used in **natural language processing (NLP)**

# EMBEDDINGS ARE DENSE REPRESENTATIONS

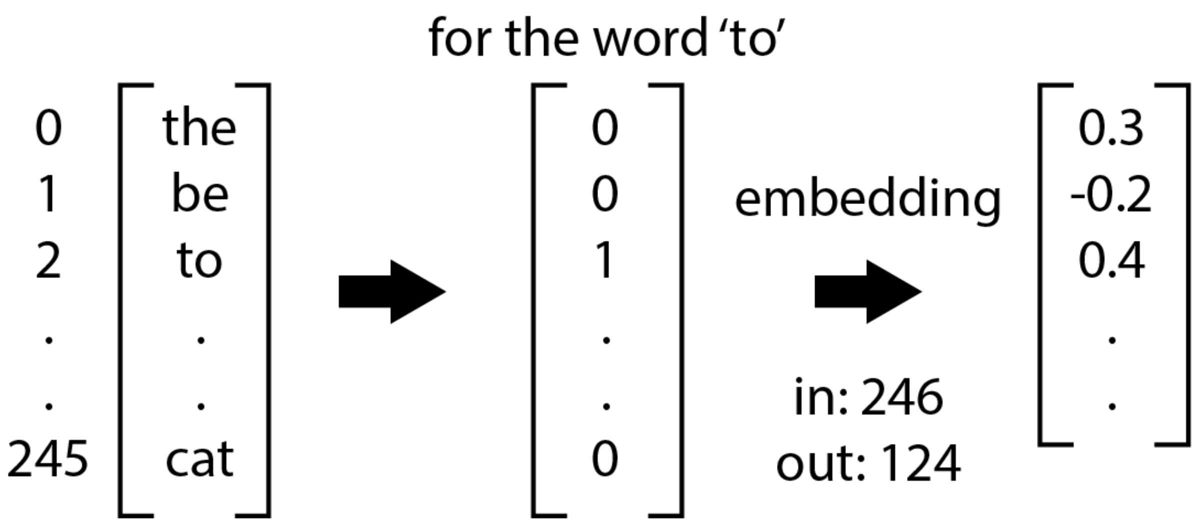



Traditionally used in **natural language processing (NLP)**
Learned **dense representations** of otherwise sparse data (vector size is a parameter)



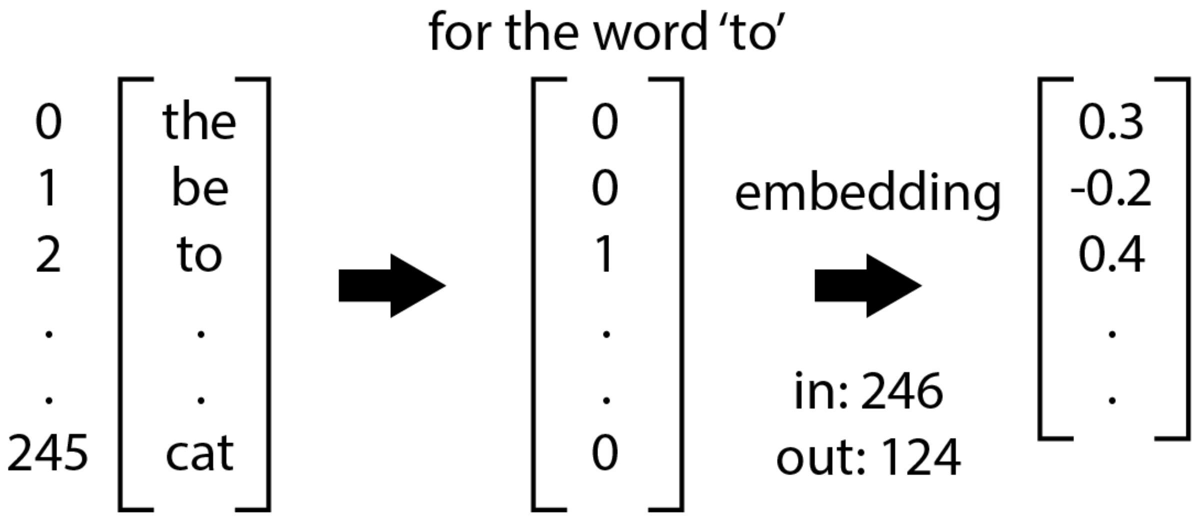
dunnhumby

# EMBEDDINGS ARE DENSE REPRESENTATIONS




Traditionally used in **natural language processing (NLP)**
Learned **dense representations** of otherwise sparse data (vector size is a parameter)
Typically learnt with **neural networks**

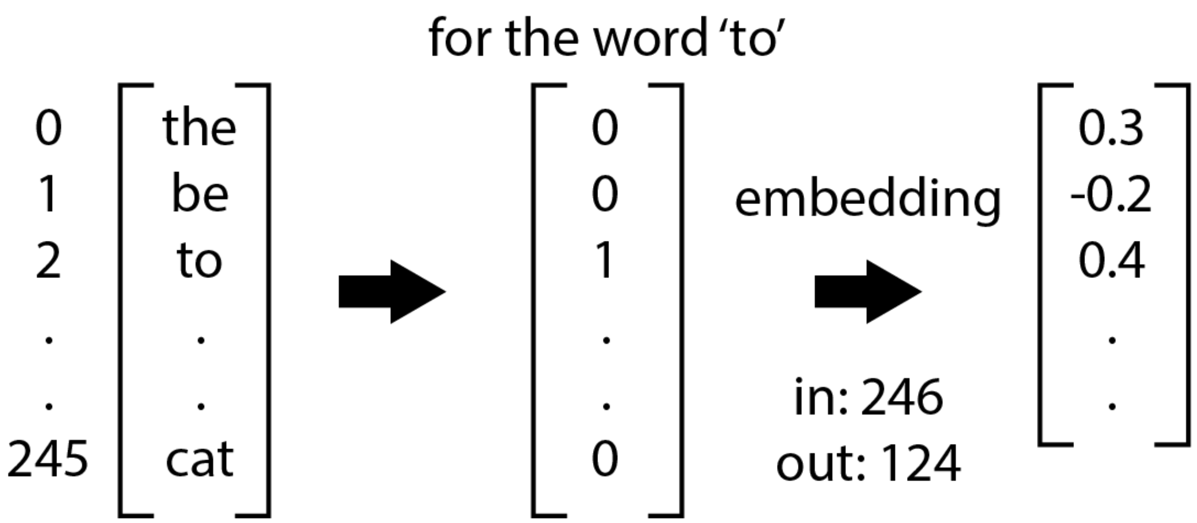
# EMBEDDINGS ARE DENSE REPRESENTATIONS




Traditionally used in **natural language processing (NLP)**
Learned **dense representations** of otherwise sparse data (vector size is a parameter)
Typically learnt with **neural networks**

When learned well, produce **similar vectors for similar items** (e.g. cats vs. dogs)

# EMBEDDINGS ARE DENSE REPRESENTATIONS

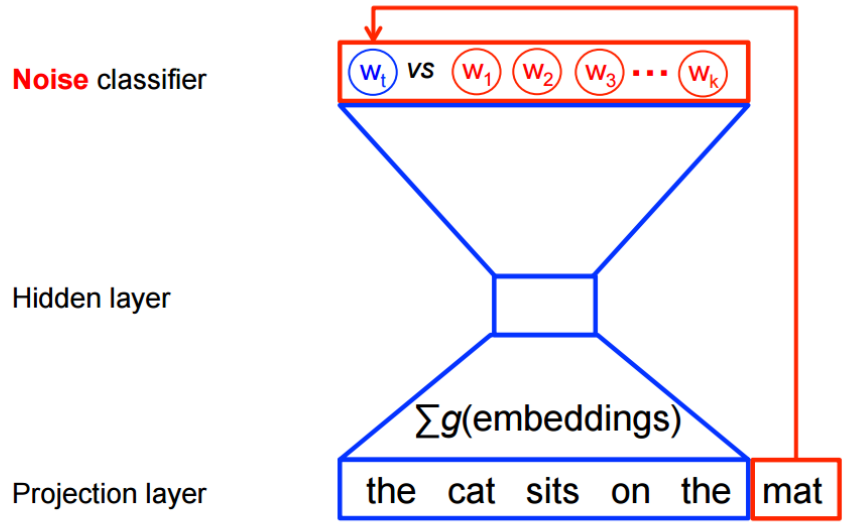



Traditionally used in **natural language processing (NLP)**
Learned **dense representations** of otherwise sparse data (vector size is a parameter)
Typically learnt with **neural networks**

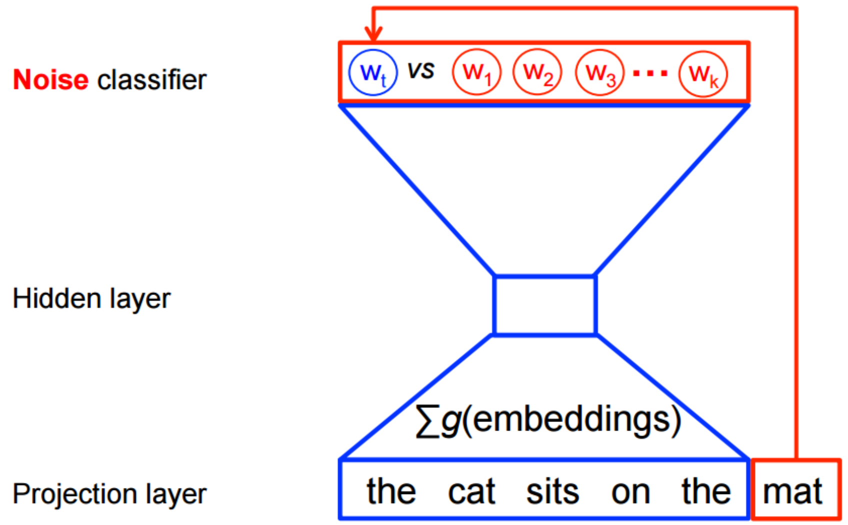
When learned well, produce **similar vectors for similar items** (e.g. cats vs. dogs)

Geoff Hinton has argued that **all thoughts can be represented** with embedding vectors

# *2VEC ALGORITHMS



# *2VEC ALGORITHMS

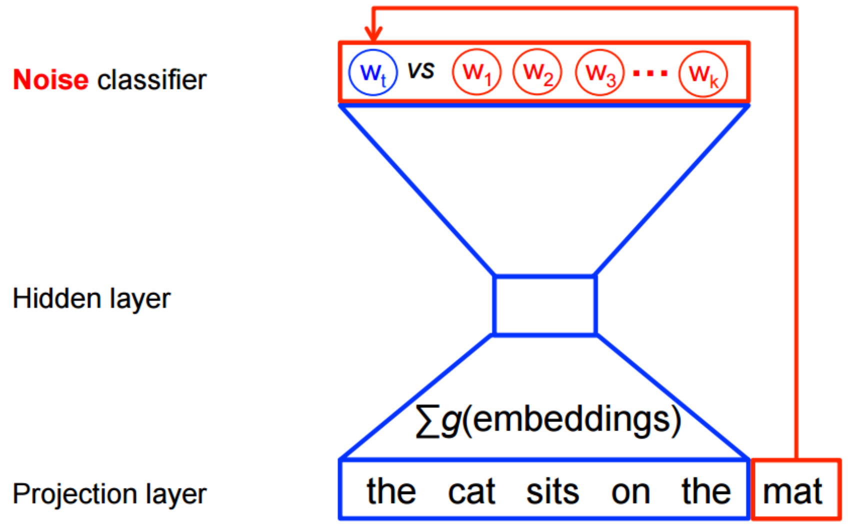



A set of **unsupervised learning** algorithms that learn vector embeddings (Mikolov et al.)



dunnhumby

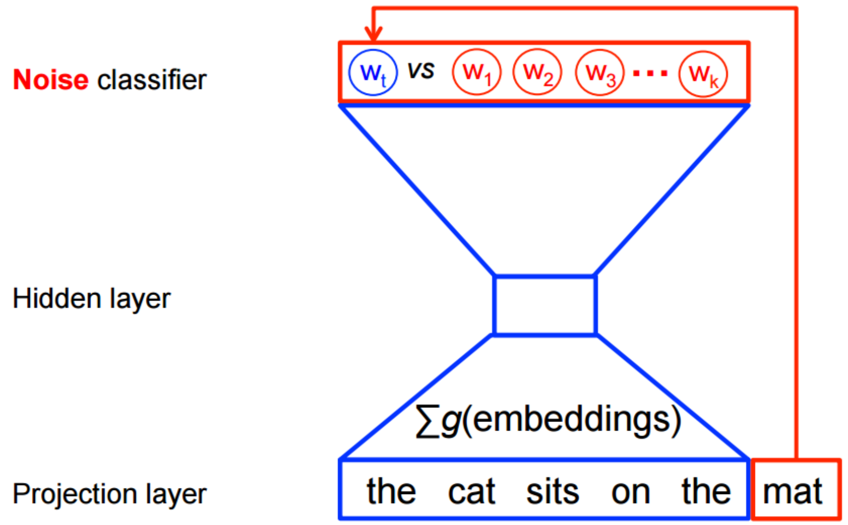
# *2VEC ALGORITHMS




A set of **unsupervised learning** algorithms that learn vector embeddings (Mikolov et al.)

**word2vec** used to learn vector embeddings for **words**

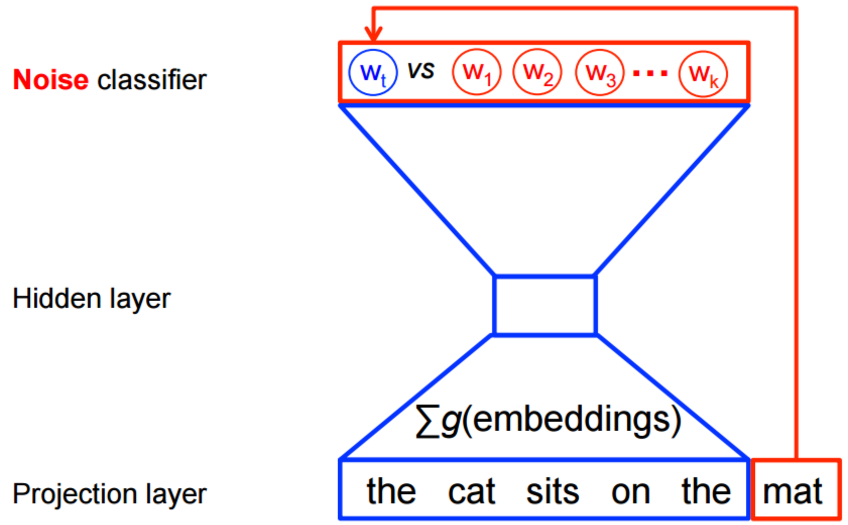
# *2VEC ALGORITHMS




A set of **unsupervised learning** algorithms that learn vector embeddings (Mikolov et al.)

**word2vec** used to learn vector embeddings for **words**
**doc2vec** used to learn vector embeddings for **documents** (e.g. sentences)

# *2VEC ALGORITHMS

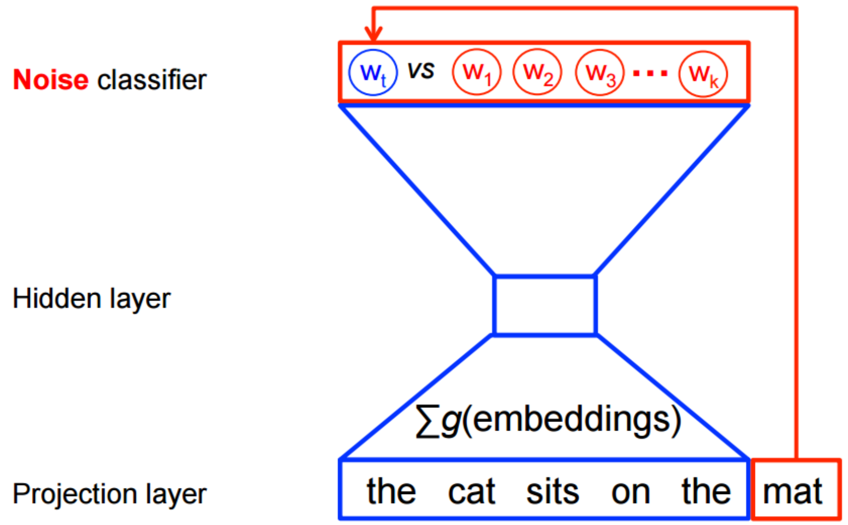



A set of **unsupervised learning** algorithms that learn vector embeddings (Mikolov et al.)

**word2vec** used to learn vector embeddings for **words**
**doc2vec** used to learn vector embeddings for **documents** (e.g. sentences)

Create **lots of little classification problems** (e.g. use vector from one word to predict next word in sentence, then backprop)

# *2VEC ALGORITHMS




A set of **unsupervised learning** algorithms that learn vector embeddings (Mikolov et al.)

**word2vec** used to learn vector embeddings for **words**
**doc2vec** used to learn vector embeddings for **documents** (e.g. sentences)

Create **lots of little classification problems** (e.g. use vector from one word to predict next word in sentence, then backprop)
Typically **very fast to train** and scale well to large data

# SENTENCES VS. BASKETS

"I love Rice Crispies, Wheat Thins and peanut butter"

=

# SENTENCES VS. BASKETS

"I love Rice Crispies, Wheat Thins and peanut butter"

=


NLP algorithms (like doc2vec) assume that you have **words (i.e. items) and sentences (i.e. documents)**

# SENTENCES VS. BASKETS

"I love Rice Crispies, Wheat Thins and peanut butter"

=

NLP algorithms (like doc2vec) assume that you have **words (i.e. items) and sentences (i.e. documents)**

**Products** are analogous to words and **baskets/customers** are analogous to sentences

# SENTENCES VS. BASKETS

"I love Rice Crispies, Wheat Thins and peanut butter"

=

NLP algorithms (like doc2vec) assume that you have **words (i.e. items) and sentences (i.e. documents)**

**Products** are analogous to words and **baskets/customers** are analogous to sentences
This data is a **neat place to test NLP algorithms** (e.g. no order, "stop words" and arguably more natural)

# SENTENCES VS. BASKETS

**“I love Rice Crispies, Wheat Thins and peanut butter”**

=


NLP algorithms (like doc2vec) assume that you have **words (i.e. items) and sentences (i.e. documents)**

**Products** are analogous to words and **baskets/customers** are analogous to sentences
This data is a **neat place to test NLP algorithms** (e.g. no order, "stop words" and arguably more natural)

e.g. A basket can be represented as a vector of product codes or descriptions

# SENTENCES VS. BASKETS

"I love Rice Crispies, Wheat Thins and peanut butter"

=


NLP algorithms (like doc2vec) assume that you have **words (i.e. items) and sentences (i.e. documents)**

**Products** are analogous to words and **baskets/customers** are analogous to sentences
This data is a **neat place to test NLP algorithms** (e.g. no order, "stop words" and arguably more natural)

e.g. A basket can be represented as a vector of product codes or descriptions

```
["Rice crispies", "Tomato ketchup", "Peanut butter"]
```

# THE MODEL

# THE MODELLING PROBLEM

dunnhumby | SOURCE FILES

# THE MODELLING PROBLEM

dunnhumby | SOURCE FILES

**Aim:** Understand whether doc2vec can learn **useful** vector representations of products and baskets

# THE MODELLING PROBLEM

dunnhumby | SOURCE FILES

**Aim:** Understand whether doc2vec can learn **useful** vector representations of products and baskets

Model developed on **free transactional** dataset from dunnhumby source files

# THE MODELLING PROBLEM

dunnhumby | SOURCE FILES

**Aim:** Understand whether doc2vec can learn **useful** vector representations of products and baskets

Model developed on **free transactional** dataset from dunnhumby source files
2 years worth of transactions from **2500 households**

# THE MODELLING PROBLEM

dunnhumby | SOURCE FILES

**Aim:** Understand whether doc2vec can learn **useful** vector representations of products and baskets

Model developed on **free transactional** dataset from dunnhumby source files
2 years worth of transactions from **2500 households**

Using gensim implementation of doc2vec (Python)

# THE MODELLING PROBLEM

dunnhumby | SOURCE FILES

**Aim:** Understand whether doc2vec can learn **useful** vector representations of products and baskets

Model developed on **free transactional** dataset from dunnhumby source files
2 years worth of transactions from **2500 households**

Using gensim implementation of doc2vec (Python)
Learn product and document vectors at same time

# THE MODELLING PROBLEM

dunnhumby | SOURCE FILES

**Aim:** Understand whether doc2vec can learn **useful** vector representations of products and baskets

Model developed on **free transactional** dataset from dunnhumby source files
2 years worth of transactions from **2500 households**

Using gensim implementation of doc2vec (Python)
Learn product and document vectors at same time

Trained on Google Cloud Platform Compute Engine (4 vCPUs and 15GB memory) in (~5 minutes per experiment)

# THE MODELLING PROBLEM

dunnhumby | SOURCE FILES

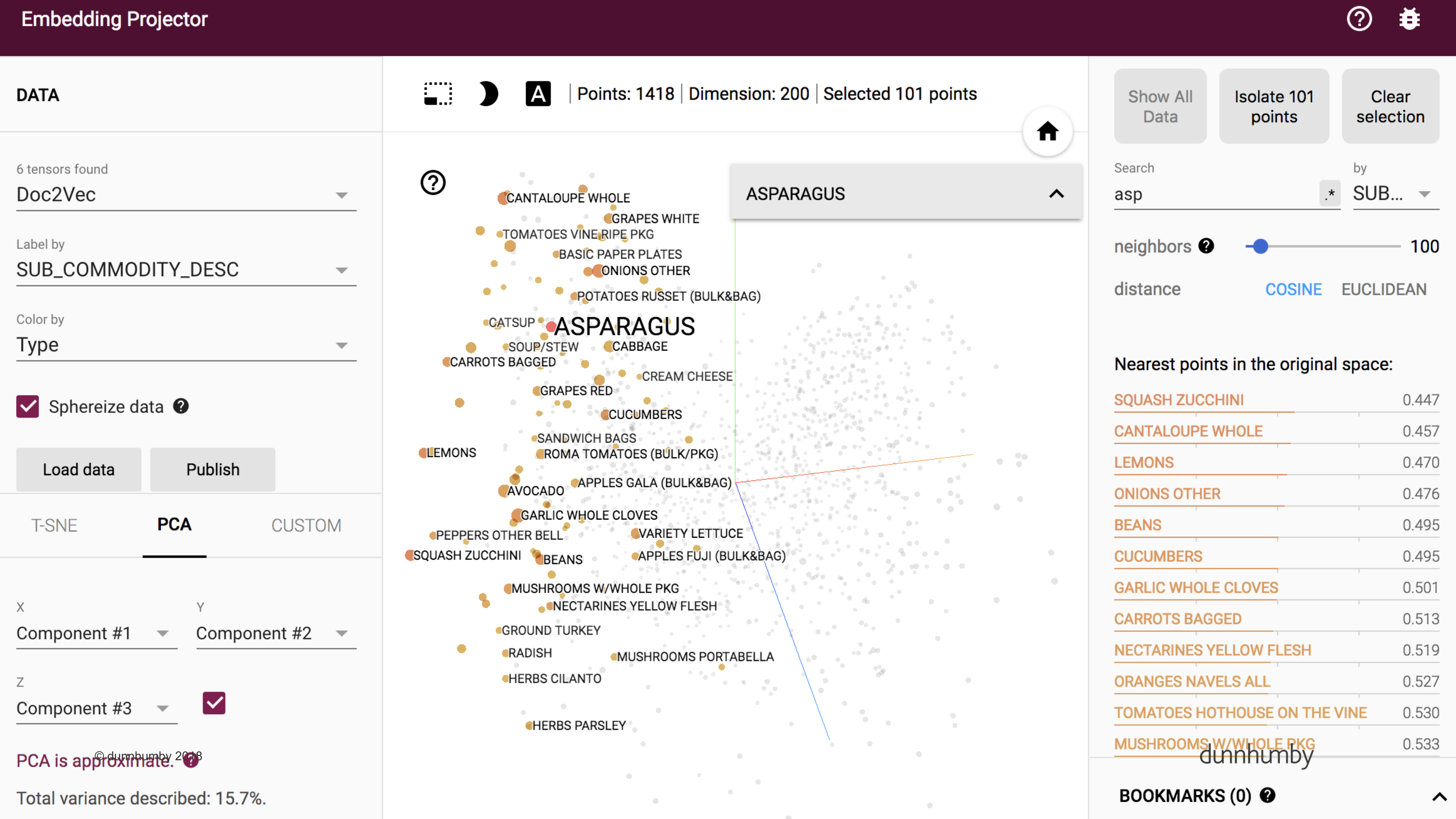
**Aim:** Understand whether doc2vec can learn **useful** vector representations of products and baskets

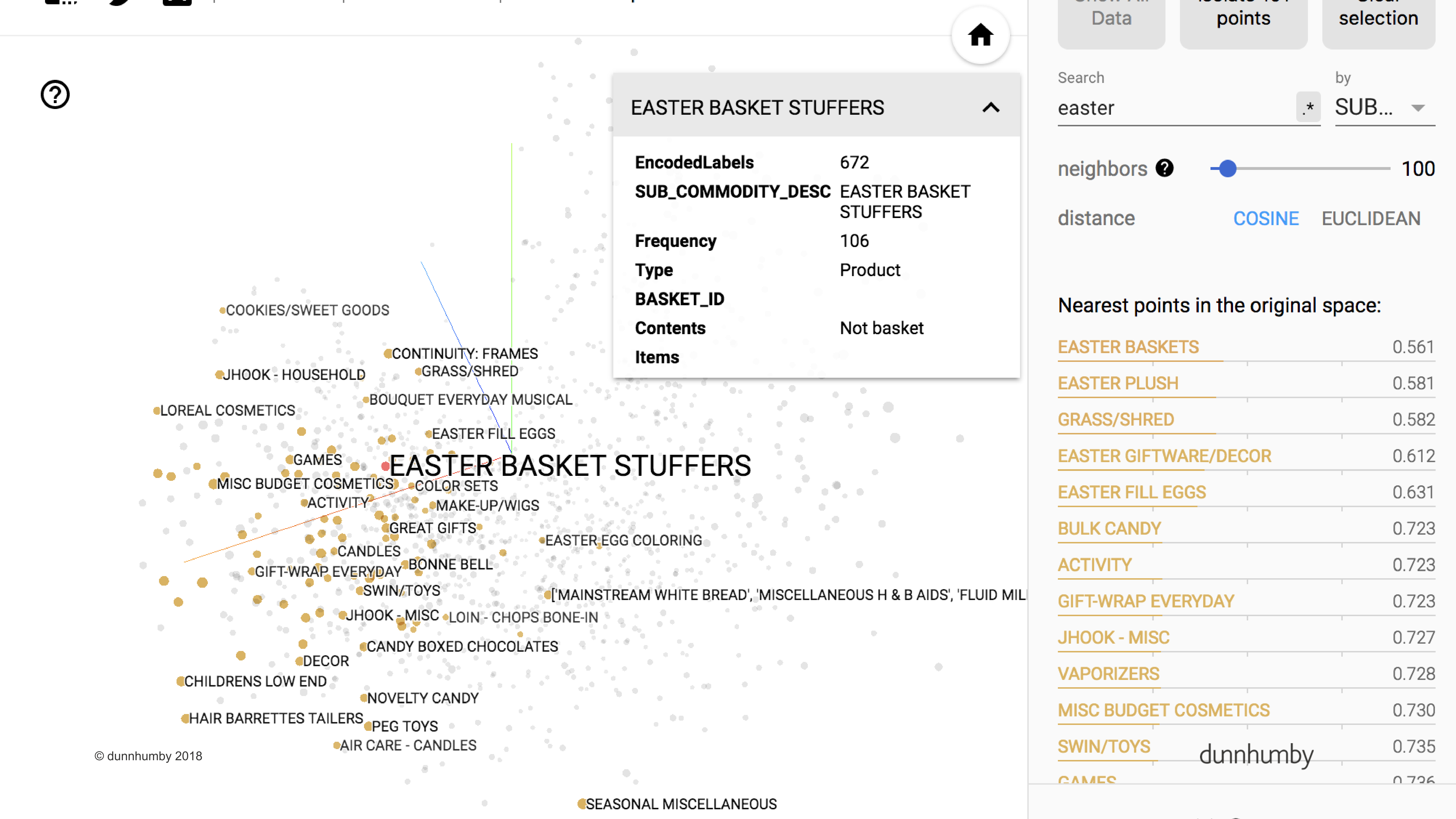
Model developed on **free transactional** dataset from dunnhumby source files
2 years worth of transactions from **2500 households**

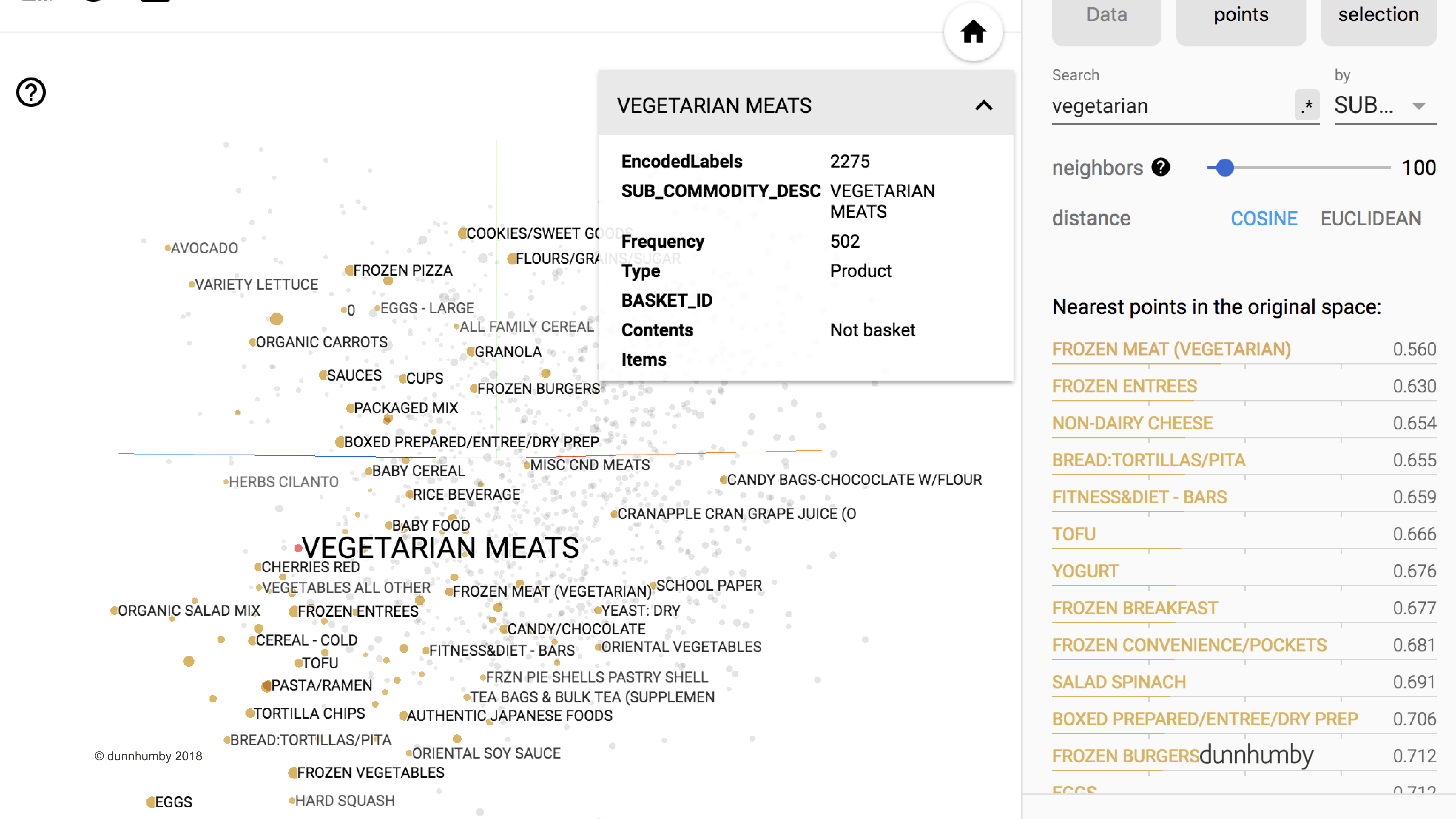
Using gensim implementation of doc2vec (Python)
Learn product and document vectors at same time

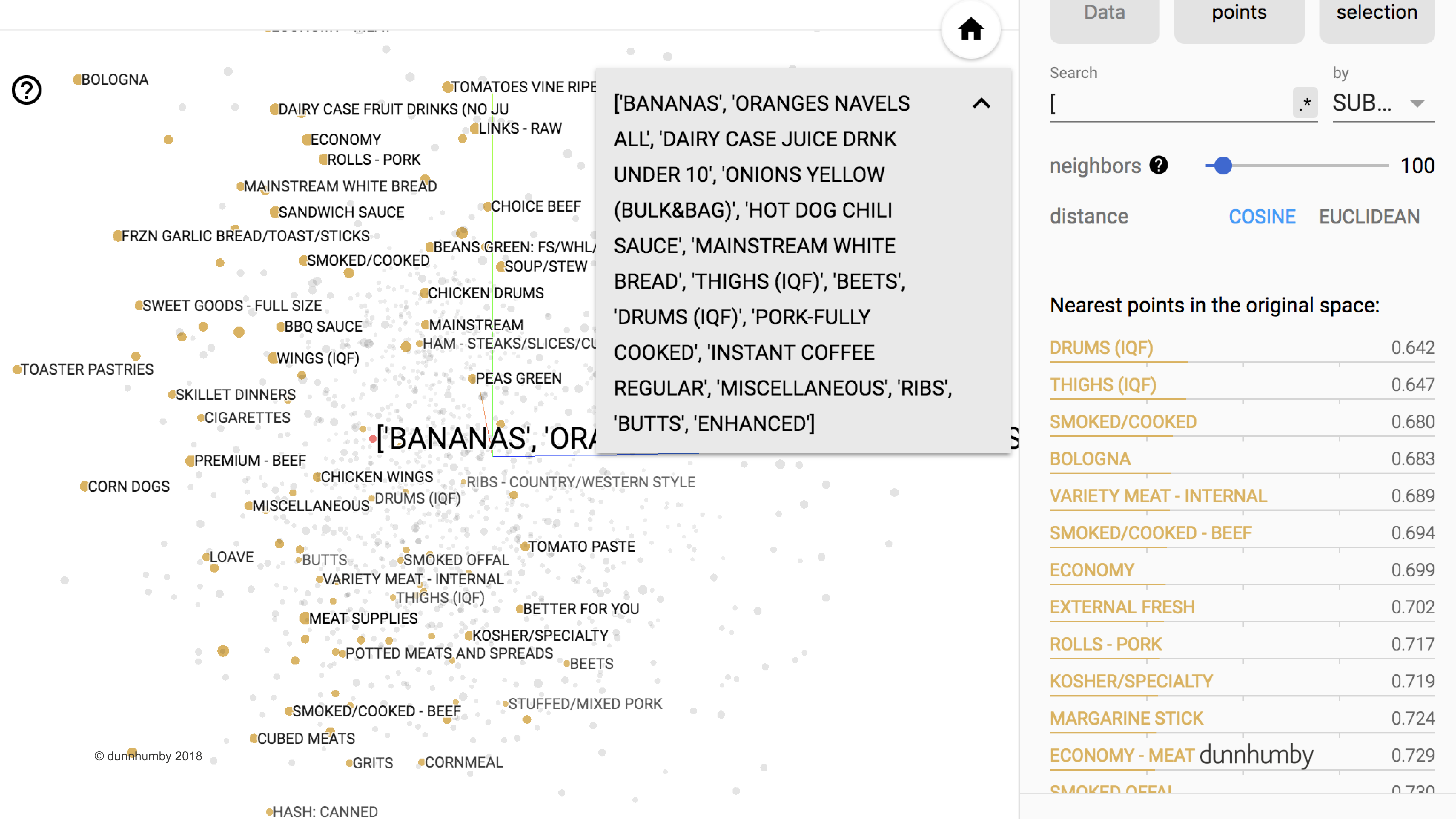
Trained on Google Cloud Platform Compute Engine (4 vCPUs and 15GB memory) in (~5 minutes per experiment)
Model solutions evaluated using Tensorflow's amazing Embedding Explorer

# EVALUATION

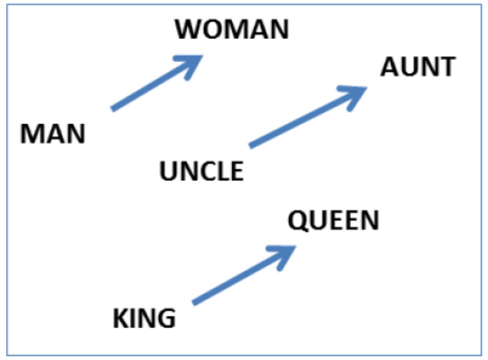

Embedding Projector
DATA
6 tensors found
Doc2Vec
Label by
SUB_COMMODITY_DESC
Color by
Type
Sphereize data
Load data
Publish
T-SNE
PCA
CUSTOM
X
Component #1
Y
Component #2
Z
Component #3
PCA is approximate.
Total variance described: 15.7%.
Points: 1418 | Dimension: 200 | Selected 101 points
ASPARAGUS
CANTALOUPE WHOLE
GRAPES WHITE
TOMATOES VINE RIPE PKG
BASIC PAPER PLATES
ONIONS OTHER
POTATOES RUSSET (BULK&BAG)
CATSUP
ASPARAGUS
SOUP/STEW
CABBAGE
CARROTS BAGGED
CREAM CHEESE
GRAPES RED
CUCUMBERS
SANDWICH BAGS
LEMONS
ROMA TOMATOES (BULK/PKG)
APPLES GALA (BULK&BAG)
AVOCADO
GARLIC WHOLE CLOVES
PEPPERS OTHER BELL
VARIETY LETTUCE
SQUASH ZUCCHINI
BEANS
APPLES FUJI (BULK&BAG)
MUSHROOMS W/WHOLE PKG
NECTARINES YELLOW FLESH
GROUND TURKEY
RADISH
MUSHROOMS PORTABELLA
HERBS CILANTO
HERBS PARSLEY
Show All Data
Isolate 101 points
Clear selection
Search
asp
by
SUB...
neighbors
100
distance
COSINE
EUCLIDEAN
Nearest points in the original space:
SQUASH ZUCCHINI 0.447
CANTALOUPE WHOLE 0.457
LEMONS 0.470
ONIONS OTHER 0.476
BEANS 0.495
CUCUMBERS 0.495
GARLIC WHOLE CLOVES 0.501
CARROTS BAGGED 0.513
NECTARINES YELLOW FLESH 0.519
ORANGES NAVELS ALL 0.527
TOMATOES HOTHOUSE ON THE VINE 0.530
MUSHROOMS W/WHOLE PKG 0.533
BOOKMARKS (0)
© dunnhumby 2018
dunnhumby

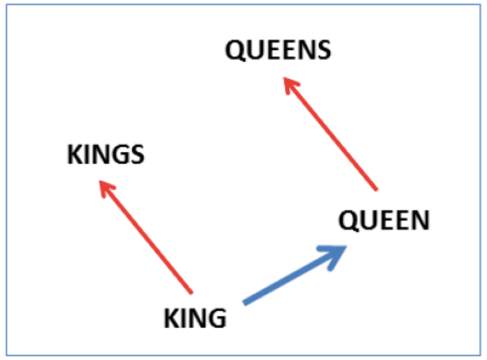

Data
points
selection
EASTER BASKET STUFFERS
EncodedLabels 672
SUB_COMMODITY_DESC EASTER BASKET STUFFERS
Frequency 106
Type Product
BASKET_ID
Contents Not basket
Items
Search
easter
by
SUB...
neighbors
100
distance
COSINE
EUCLIDEAN
Nearest points in the original space:
EASTER BASKETS 0.561
EASTER PLUSH 0.581
GRASS/SHRED 0.582
EASTER GIFTWARE/DECOR 0.612
EASTER FILL EGGS 0.631
BULK CANDY 0.723
ACTIVITY 0.723
GIFT-WRAP EVERYDAY 0.723
JHOOK - MISC 0.727
VAPORIZERS 0.728
MISC BUDGET COSMETICS 0.730
SWIN/TOYS 0.735
dunnhumby
COOKIES/SWEET GOODS
CONTINUITY: FRAMES
JHOOK - HOUSEHOLD
GRASS/SHRED
BOUQUET EVERYDAY MUSICAL
LOREAL COSMETICS
EASTER FILL EGGS
GAMES
EASTER BASKET STUFFERS
MISC BUDGET COSMETICS
COLOR SETS
ACTIVITY
MAKE-UP/WIGS
GREAT GIFTS
EASTER EGG COLORING
CANDLES
BONNE BELL
GIFT-WRAP EVERYDAY
SWIN/TOYS
['MAINSTREAM WHITE BREAD', 'MISCELLANEOUS H & B AIDS', 'FLUID MIL
JHOOK - MISC
LOIN - CHOPS BONE-IN
CANDY BOXED CHOCOLATES
DECOR
CHILDRENS LOW END
NOVELTY CANDY
HAIR BARRETTES TAILERS
PEG TOYS
AIR CARE - CANDLES
© dunnhumby 2018
SEASONAL MISCELLANEOUS


Data
points
selection
Search
by
vegetarian
SUB...
neighbors
100
distance
COSINE
EUCLIDEAN
Nearest points in the original space:
FROZEN MEAT (VEGETARIAN) 0.560
FROZEN ENTREES 0.630
NON-DAIRY CHEESE 0.654
BREAD:TORTILLAS/PITA 0.655
FITNESS&DIET - BARS 0.659
TOFU 0.666
YOGURT 0.676
FROZEN BREAKFAST 0.677
FROZEN CONVENIENCE/POCKETS 0.681
SALAD SPINACH 0.691
BOXED PREPARED/ENTREE/DRY PREP 0.706
FROZEN BURGERS 0.712
dunnhumby
VEGETARIAN MEATS
EncodedLabels 2275
SUB_COMMODITY_DESC VEGETARIAN MEATS
Frequency 502
Type Product
BASKET_ID
Contents Not basket
Items
AVOCADO
VARIETY LETTUCE
FROZEN PIZZA
COOKIES/SWEET GOODS
FLOURS/GRAINS/SUGAR
0
EGGS - LARGE
ALL FAMILY CEREAL
ORGANIC CARROTS
GRANOLA
SAUCES
CUPS
FROZEN BURGERS
PACKAGED MIX
BOXED PREPARED/ENTREE/DRY PREP
BABY CEREAL
MISC CND MEATS
HERBS CILANTO
CANDY BAGS-CHOCOCLATE W/FLOUR
RICE BEVERAGE
CRANAPPLE CRAN GRAPE JUICE (O
BABY FOOD
VEGETARIAN MEATS
CHERRIES RED
VEGETABLES ALL OTHER
FROZEN MEAT (VEGETARIAN)
SCHOOL PAPER
ORGANIC SALAD MIX
FROZEN ENTREES
YEAST: DRY
CANDY/CHOCOLATE
CEREAL - COLD
FITNESS&DIET - BARS
ORIENTAL VEGETABLES
TOFU
FRZN PIE SHELLS PASTRY SHELL
PASTA/RAMEN
TEA BAGS & BULK TEA (SUPPLEMEN
TORTILLA CHIPS
AUTHENTIC JAPANESE FOODS
BREAD:TORTILLAS/PITA
© dunnhumby 2018
ORIENTAL SOY SAUCE
FROZEN VEGETABLES
EGGS
HARD SQUASH

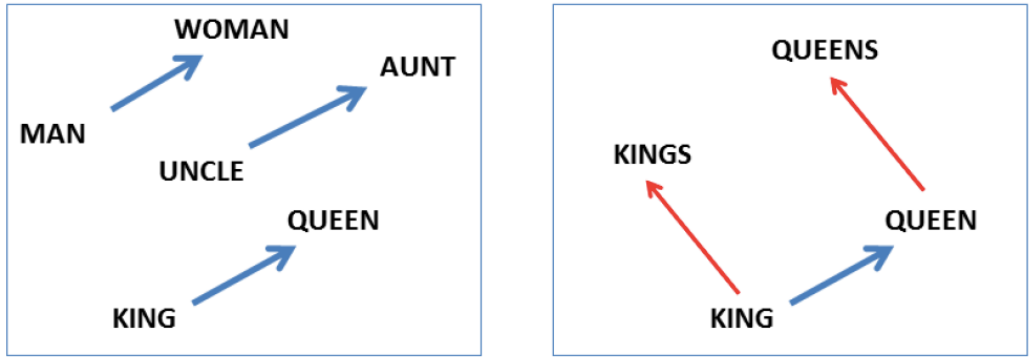

BOLOGNA
TOMATOES VINE RIPE
DAIRY CASE FRUIT DRINKS (NO JU
LINKS - RAW
ECONOMY
ROLLS - PORK
MAINSTREAM WHITE BREAD
CHOICE BEEF
SANDWICH SAUCE
FRZN GARLIC BREAD/TOAST/STICKS
BEANS GREEN: FS/WHL/
SMOKED/COOKED
SOUP/STEW
CHICKEN DRUMS
SWEET GOODS - FULL SIZE
BBQ SAUCE
MAINSTREAM
HAM - STEAKS/SLICES/CU
WINGS (IQF)
TOASTER PASTRIES
PEAS GREEN
SKILLET DINNERS
CIGARETTES
['BANANAS', 'ORA
PREMIUM - BEEF
CHICKEN WINGS
RIBS - COUNTRY/WESTERN STYLE
CORN DOGS
MISCELLANEOUS
DRUMS (IQF)
TOMATO PASTE
LOAVE
BUTTS
SMOKED OFFAL
VARIETY MEAT - INTERNAL
THIGHS (IQF)
BETTER FOR YOU
MEAT SUPPLIES
KOSHER/SPECIALTY
POTTED MEATS AND SPREADS
BEETS
SMOKED/COOKED - BEEF
STUFFED/MIXED PORK
CUBED MEATS
© dunnhumby 2018
GRITS
CORNMEAL
HASH: CANNED
['BANANAS', 'ORANGES NAVELS ALL', 'DAIRY CASE JUICE DRNK UNDER 10', 'ONIONS YELLOW (BULK&BAG)', 'HOT DOG CHILI SAUCE', 'MAINSTREAM WHITE BREAD', 'THIGHS (IQF)', 'BEETS', 'DRUMS (IQF)', 'PORK-FULLY COOKED', 'INSTANT COFFEE REGULAR', 'MISCELLANEOUS', 'RIBS', 'BUTTS', 'ENHANCED']
Data
points
selection
Search
by
[
.*
SUB...
neighbors
100
distance
COSINE
EUCLIDEAN
Nearest points in the original space:
DRUMS (IQF) 0.642
THIGHS (IQF) 0.647
SMOKED/COOKED 0.680
BOLOGNA 0.683
VARIETY MEAT - INTERNAL 0.689
SMOKED/COOKED - BEEF 0.694
ECONOMY 0.699
EXTERNAL FRESH 0.702
ROLLS - PORK 0.717
KOSHER/SPECIALTY 0.719
MARGARINE STICK 0.724
ECONOMY - MEAT 0.729
dunnhumby

# ANALOGICAL REASONING



# ANALOGICAL REASONING

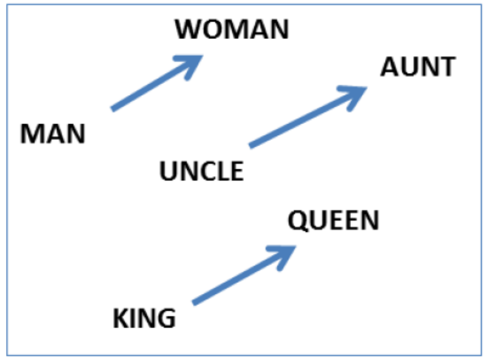


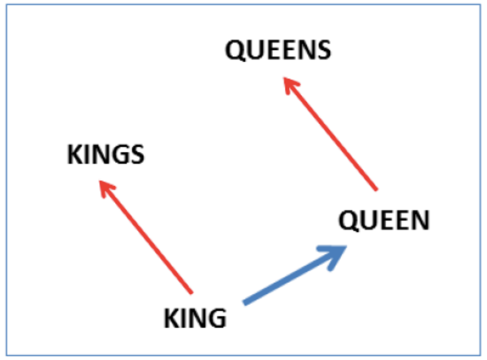

**Semantic relationships between words** are typically preserved within embedding space



dunnhumby

# ANALOGICAL REASONING

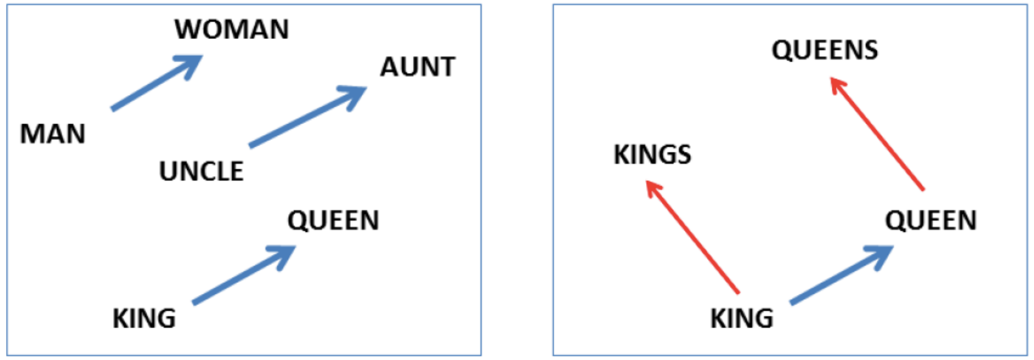



**Semantic relationships between words** are typically preserved within embedding space

$$King - Man + Woman = y$$

# ANALOGICAL REASONING

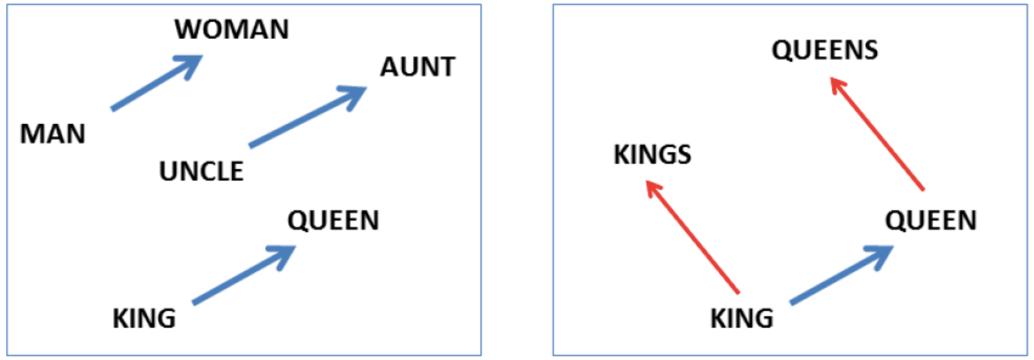



**Semantic relationships between words** are typically preserved within embedding space

$$King - Man + Woman = y$$

Vectors most similar to $y$ tend to be **related to "queen"**

# ANALOGICAL REASONING




**Semantic relationships between words** are typically preserved within embedding space

$$King - Man + Woman = y$$

Vectors most similar to $y$ tend to be **related to "queen"**

What would be the equivalent using **product embeddings**?

# IDENTIFYING CHEAPER ALTERNATIVES

**Premium Meat - Premium + Economy = ?**

# IDENTIFYING CHEAPER ALTERNATIVES

**Premium Meat - Premium + Economy = ?**

| Rank | Product description | Cosine similarity |
|---|---|---|
| **1** | Economy meat | 0.542439 |
| **2** | Margarine stick | 0.518072 |
| **3** | Carrots (bagged) | 0.500584 |
| **4** | Chocolate milk | 0.489882 |

# IDENTIFYING CHEAPER ALTERNATIVES

**Premium Meat - Premium + Economy = ?**

| Rank | Product description | Cosine similarity |
|---|---|---|
| **1** | Economy meat | 0.542439 |
| **2** | Margarine stick | 0.518072 |
| **3** | Carrots (bagged) | 0.500584 |
| **4** | Chocolate milk | 0.489882 |

The most similar product to the target vector is **economy meat**. Other items are also cheaper alternatives.

# IDENTIFYING CHEAPER ALTERNATIVES

**Premium Meat - Premium + Economy = ?**

| Rank | Product description | Cosine similarity |
|---|---|---|
| **1** | Economy meat | 0.542439 |
| **2** | Margarine stick | 0.518072 |
| **3** | Carrots (bagged) | 0.500584 |
| **4** | Chocolate milk | 0.489882 |

The most similar product to the target vector is **economy meat.** Other items are also cheaper alternatives.
Suggests that there is a **"price sensitivity"** direction within the space

# IDENTIFYING VEGETARIAN ALTERNATIVES

Frozen Burgers - Beef + Tofu = ?

# IDENTIFYING VEGETARIAN ALTERNATIVES

**Frozen Burgers - Beef + Tofu = ?**

| Rank | Product description | Cosine similarity |
|---|---|---|
| **1** | Frozen entrees | 0.518292 |
| **2** | Frozen meat (vegetarian) | 0.484718 |
| **3** | Frozen meal combo/dinners | 0.464290 |
| **4** | Eggs | 0.462903 |

# IDENTIFYING VEGETARIAN ALTERNATIVES

**Frozen Burgers - Beef + Tofu = ?**

| Rank | Product description | Cosine similarity |
|---|---|---|
| **1** | Frozen entrees | 0.518292 |
| **2** | Frozen meat (vegetarian) | 0.484718 |
| **3** | Frozen meal combo/dinners | 0.464290 |
| **4** | Eggs | 0.462903 |

The most similar products to the target vector tend to be **frozen and/or meat free**

# IDENTIFYING VEGETARIAN ALTERNATIVES

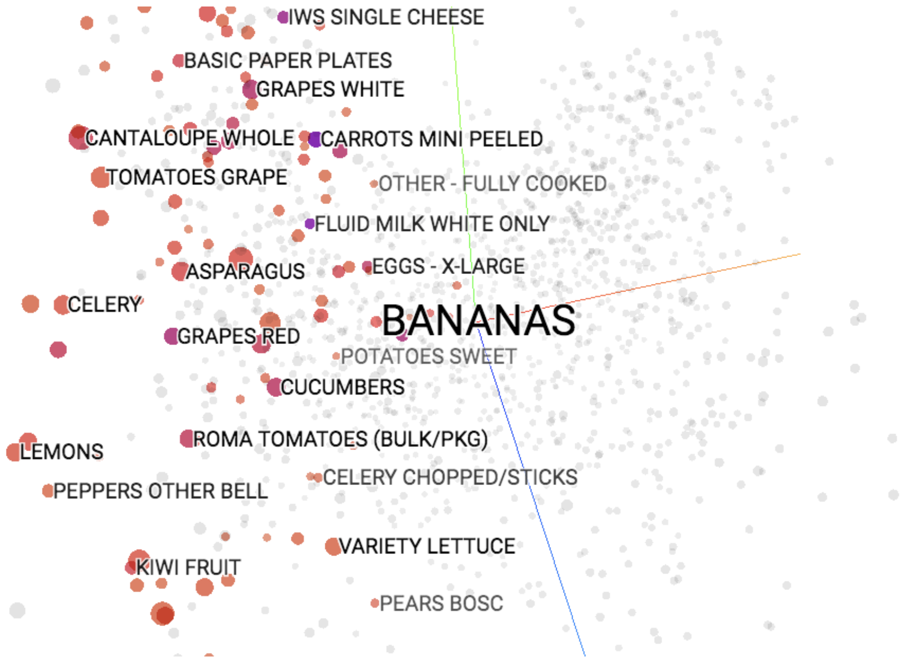
**Frozen Burgers - Beef + Tofu = ?**

| Rank | Product description | Cosine similarity |
|---|---|---|
| **1** | Frozen entrees | 0.518292 |
| **2** | Frozen meat (vegetarian) | 0.484718 |
| **3** | Frozen meal combo/dinners | 0.464290 |
| **4** | Eggs | 0.462903 |

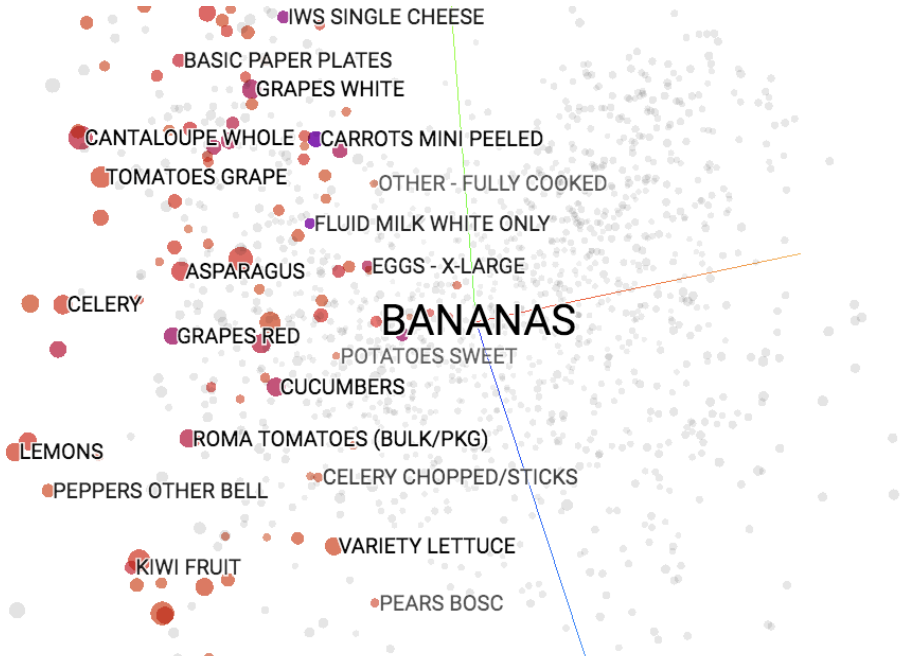
The most similar products to the target vector tend to be **frozen and/or meat free**
Suggests there are **"frozen"** and **"vegetarian"** directions within the embedding space

# CONCLUSIONS

# THE MODEL LOOKS GREAT

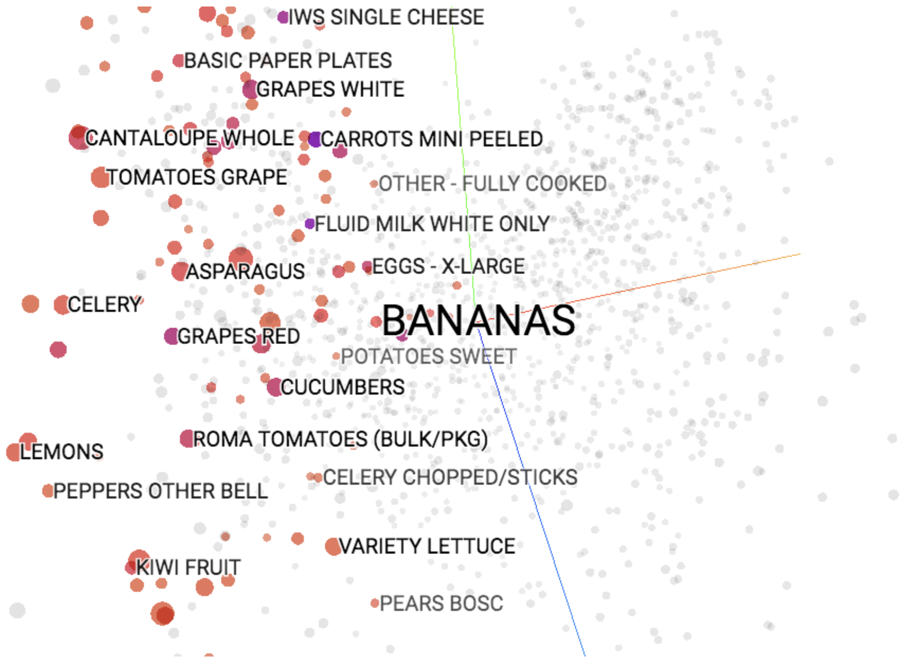


# THE MODEL LOOKS GREAT




+ Very **visual** application of machine learning in retail

# THE MODEL LOOKS GREAT

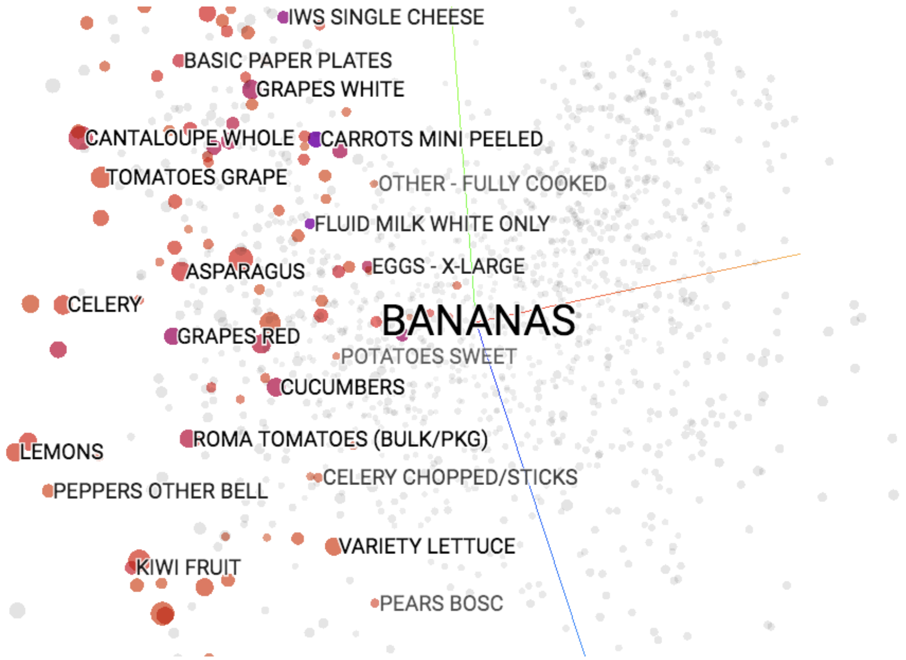



+ Very **visual** application of machine learning in retail
+ The model appears to **understand product similarity** well



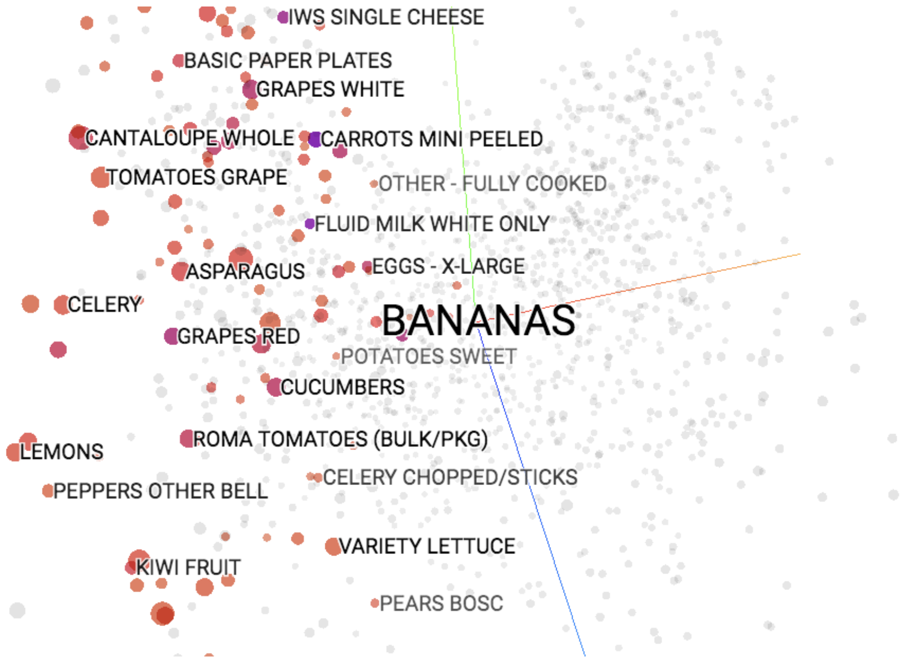
dunnhumby

# THE MODEL LOOKS GREAT



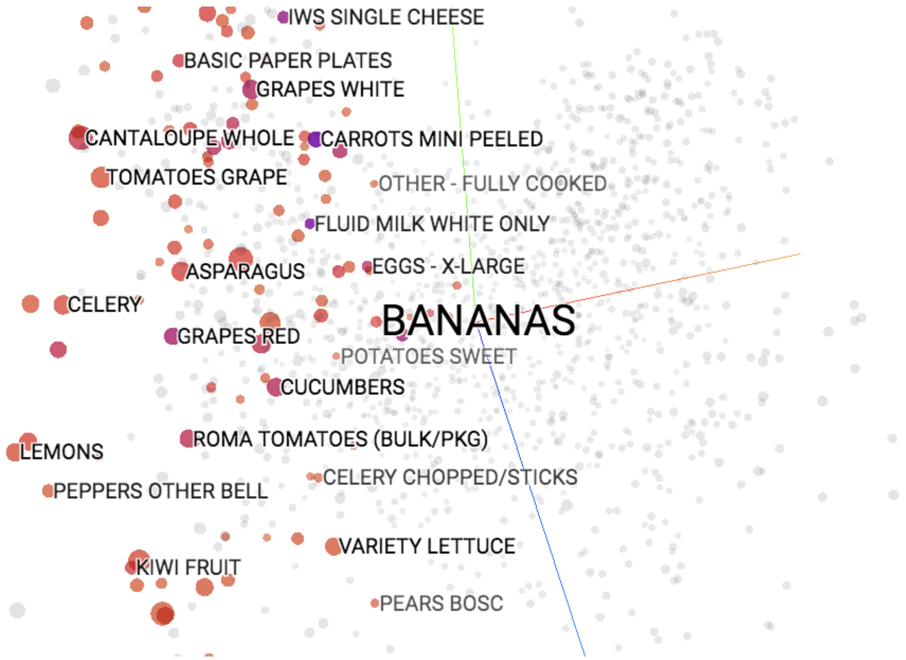

+ Very **visual** application of machine learning in retail
+ The model appears to **understand product similarity** well
+ **Analogical reasoning** results seem intuitive

# THE MODEL LOOKS GREAT




+ Very **visual** application of machine learning in retail
+ The model appears to **understand product similarity** well
+ **Analogical reasoning** results seem intuitive
+ **Scales** better than rival algorithms (e.g. NMF)

# THE MODEL LOOKS GREAT



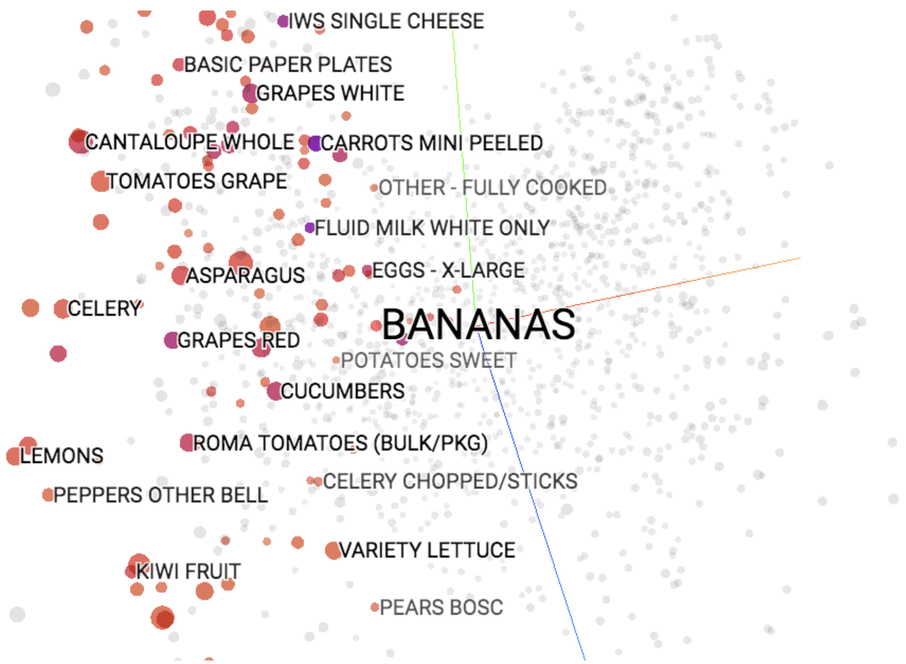

+ Very **visual** application of machine learning in retail
+ The model appears to **understand product similarity** well
+ **Analogical reasoning** results seem intuitive
+ **Scales** better than rival algorithms (e.g. NMF)

- Harder with **very low velocity** items



dunnhumby

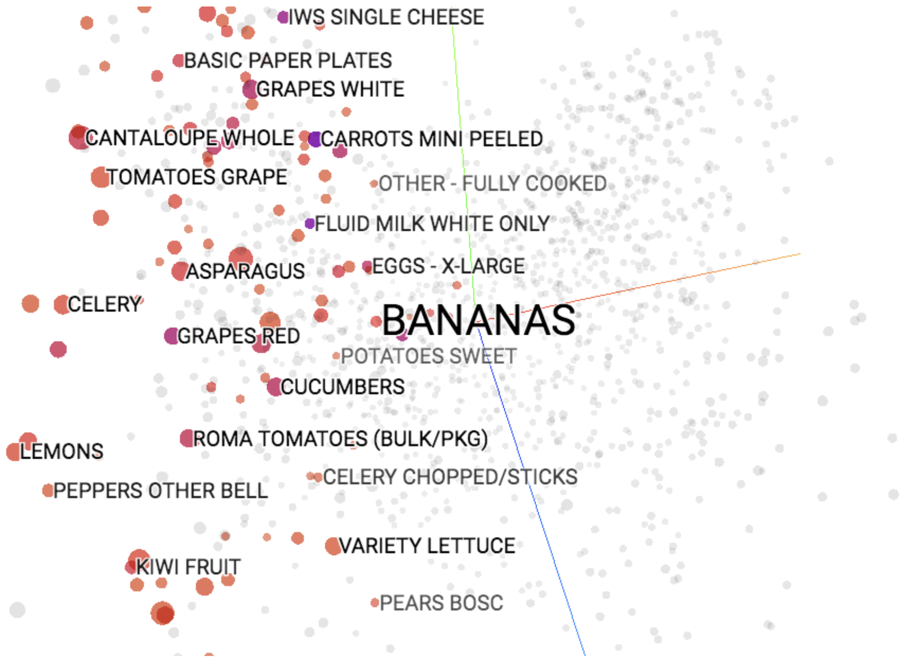
# THE MODEL LOOKS GREAT




+ Very **visual** application of machine learning in retail
+ The model appears to **understand product similarity** well
+ **Analogical reasoning** results seem intuitive
+ **Scales** better than rival algorithms (e.g. NMF)

- Harder with **very low velocity** items
- Need some more **automated** ways of **evaluating** the model

# THE MODEL LOOKS GREAT



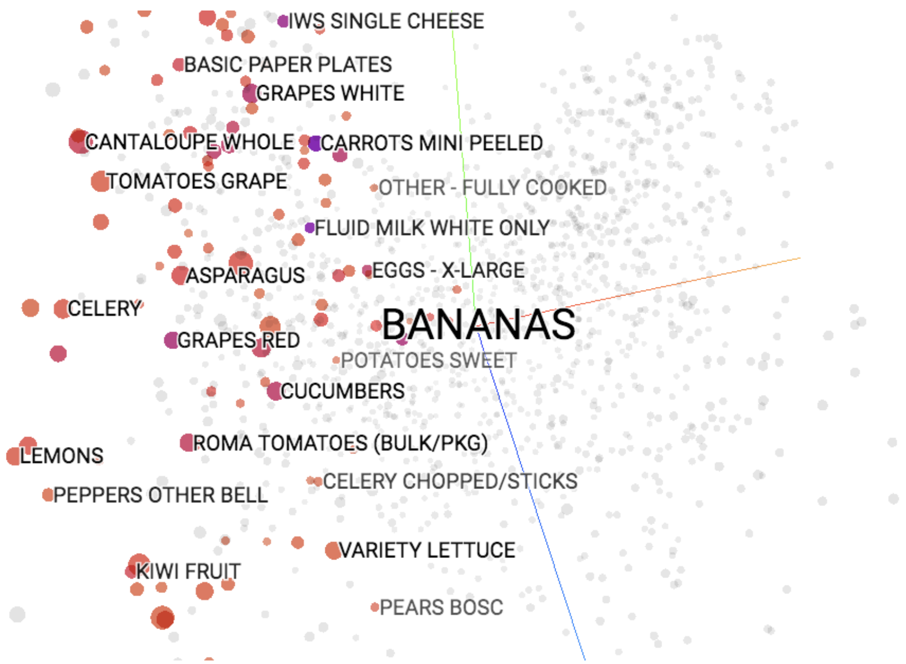

+ Very **visual** application of machine learning in retail
+ The model appears to **understand product similarity** well
+ **Analogical reasoning** results seem intuitive
+ **Scales** better than rival algorithms (e.g. NMF)

- Harder with **very low velocity** items
- Need some more **automated** ways of **evaluating** the model
- Need **less homogenous** way of representing baskets & customers



dunnhumby

# THE MODEL LOOKS GREAT




+ Very **visual** application of machine learning in retail
+ The model appears to **understand product similarity** well
+ **Analogical reasoning** results seem intuitive
+ **Scales** better than rival algorithms (e.g. NMF)

- Harder with **very low velocity** items
- Need some more **automated** ways of **evaluating** the model
- Need **less homogenous** way of representing baskets & customers
- **Interpretation of "similar"** is slightly unusual (neither comps nor subs)

# THIS MODEL HAS MANY APPLICATIONS

# THIS MODEL HAS MANY APPLICATIONS

**Global deployment** of model in collaboration with partners

# THIS MODEL HAS MANY APPLICATIONS

**Global deployment** of model in collaboration with partners

**Ultra-fast deployment** of basket and customer **segmentations** with new clients

# THIS MODEL HAS MANY APPLICATIONS

**Global deployment** of model in collaboration with partners

**Ultra-fast deployment** of basket and customer **segmentations** with new clients
Exploring use of model to suggest **alternative & supplementary** products in ranges during **store ranging**



dunnhumby

# THIS MODEL HAS MANY APPLICATIONS

**Global deployment** of model in collaboration with partners

**Ultra-fast deployment** of basket and customer **segmentations** with new clients
Exploring use of model to suggest **alternative & supplementary** products in ranges during **store ranging**
Exploring power of new features in **personalised recommender systems**

# THIS MODEL HAS MANY APPLICATIONS

**Global deployment** of model in collaboration with partners

**Ultra-fast deployment** of basket and customer **segmentations** with new clients
Exploring use of model to suggest **alternative & supplementary** products in ranges during **store ranging**
Exploring power of new features in **personalised recommender systems**

A **significant leap forward** in the endeavour to describe **similarity**



dunnhumby

# THANK YOU

@adamnhornsby
adam.hornsby@dunnhumby.com

**dunnhumby.com/careers**