

# Data Analysis and Visualization with MongoDB

Alexander C. S. Hendorf

 @hendorf

MongoDB World 2016, NYC



## Alexander C. S. Hendorf

CTO Königsweg GmbH

mongoDB master 2016, MUG Leader

EuroPython organizer + program chair

Speaker EuroPython, mongoDB days, CEBIT, PyCon It, PyData...

Hobbies: see above

 @hendorf



# EUROPYTHON

## 2016

Bilbao, 17-24 July

#15



beginners' day

20 free trainings

5d talks & trainings

180+ sessions



panels

social event

open spaces

interactive sessions

2d sprints

17th - 24th of July  
@EuroPython





mongoDB<sup>®</sup>

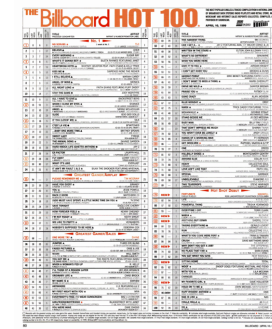
FOR **GIANT** IDEAS

2003

2012

2016





2003

2012

2016



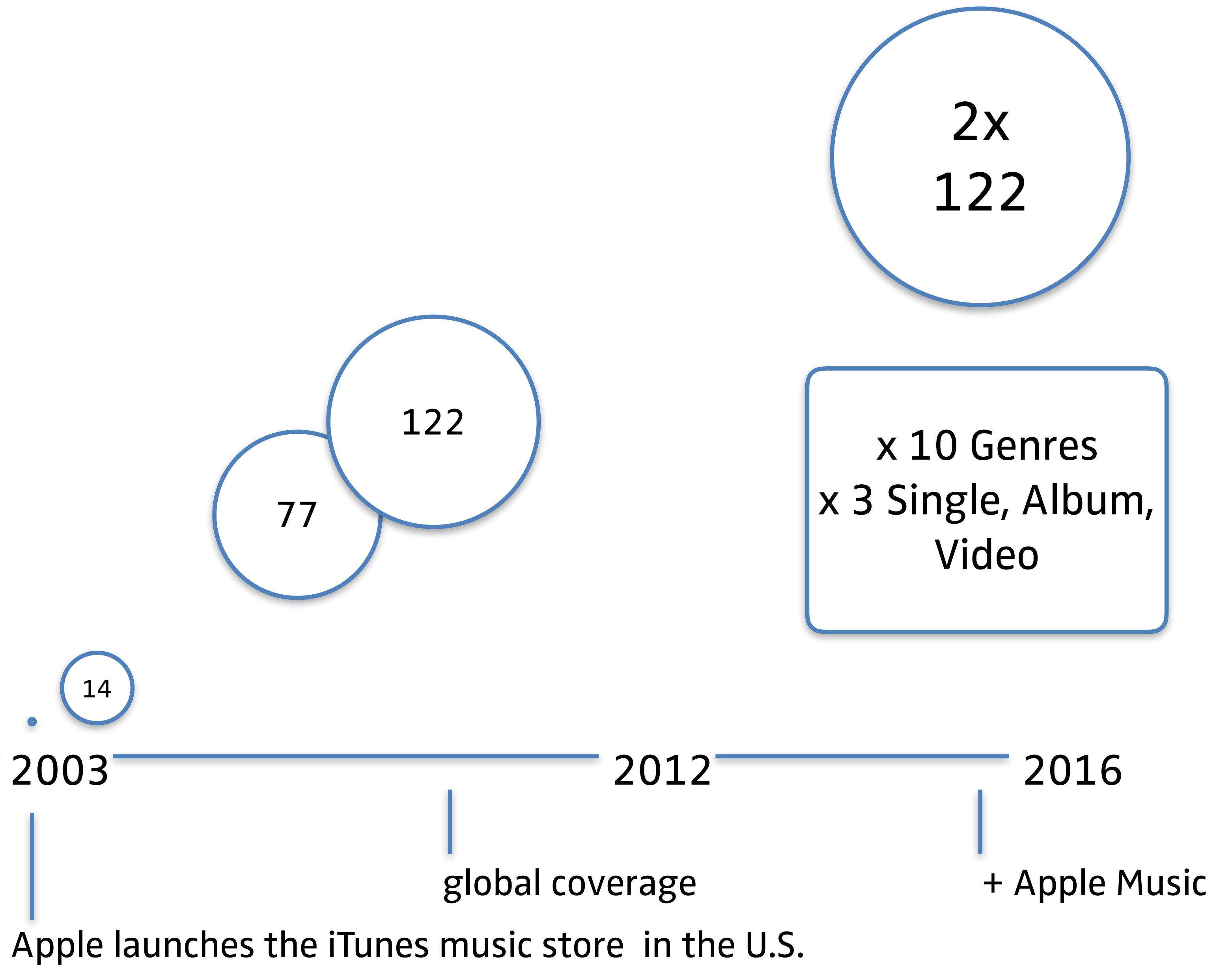
2003

2012

2016

Apple launches the iTunes music store in the U.S.









CHARTGEIST

# ***RDBS Data Lake***



# *mongoDB Data Lake*



# ***Aggregation Framework***








# Topalben


Songs Alben Musikvideos


Electronic


- 

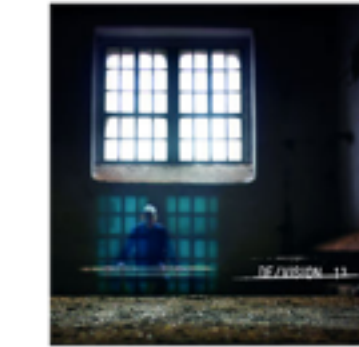
1. Summer Sessions 2016 (Compiled and Mixed by Various Artists)  
9,99 €
- 


2. The Mountain Will Fall  
DJ Shadow  
8,99 €
- 


3. Milchbar: Seaside Season 8 (Deluxe Edition)  
Blank & Jones  
9,99 €
- 

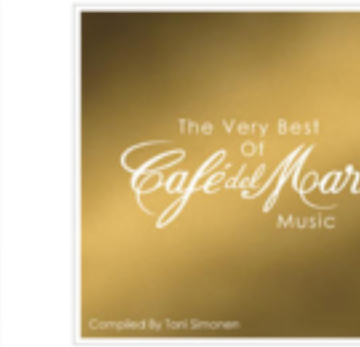
4. This Girl - EP  
Kungs  
3,49 €
- 


5. Nassau Beach Club Ibiza 2016  
Various Artists  
9,99 €
- 


6. Déepalma Ibiza 2016 (Mixed by Various Artists)  
Various Artists  
8,99 €
- 


7. 13 (Bonus Edition)  
De/Vision  
9,99 €
- 


8. Skin  
Flume  
10,99 €
- 


9. 7  
Paul Kalkbrenner  
9,99 €
- 


10. The Very Best of Cafe del Mar Music  
Café del Mar  
11,99 €
- 


11. Electronica 2: The Heart of Noise  
Jean-Michel Jarre  
10,99 €
- 


12. The Colour in Anything  
James Blake  
10,99 €
- 

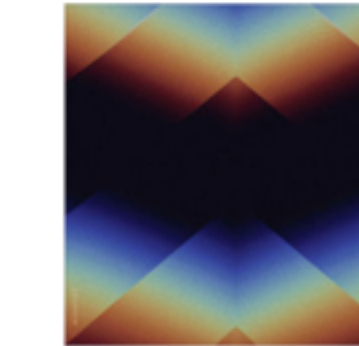
13. 10 Years Stil vor Talent  
Oliver Koletzki  
9,99 €
- 


14. Weltreise  
Schiller  
8,99 €
- 


15. Go - The Very Best of Moby...  
Moby  
13,99 €
- 


16. Built on Glass  
Chet Faker  
9,99 €
- 


17. Die Mensch-Maschine...  
Kraftwerk  
6,99 €
- 


18. Peace Is the Mission (Extended)  
Major Lazer  
8,99 €
- 


19. Weval  
Weval  
8,99 €
- 


20. Faze DJ Set #40: Stephan Bodzin  
Stephan Bodzin  
6,99 €
- 


21. The Best of Cafe del Mar - The Energy 52  
Energy 52  
2,99 €
- 


22. Milchbar - Seaside Season 6...  
Blank & Jones  
8,99 €
- 


23. Milchbar Seaside Season 1  
Blank & Jones  
7,89 €
- 


24. The Wanderings of the Avener  
The Avener  
10,99 €
- 


25. Milchbar - Seaside Season 7...  
Blank & Jones  
9,99 €
- 


26. The Art of Sampling (Special Edition)  
Parovoz  
10,99 €
- 


27. about: berlin, vol. 12  
Various Artists  
11,99 €
- 


28. Heligoland  
Massive Attack  
7,99 €
- 


29. III (Deluxe Edition)  
Moderat  
11,99 €
- 


30. The Demon Diaries (Deluxe Edition)  
Parovoz  
11,99 €
- 


31. Star of Love (Acoustic Bonus)  
Crystal Fighters  
9,99 €
- 


32. Nassau Beach Club Ibiza 2015  
Various Artists  
9,99 €
- 


33. Relax Edition 9  
Blank & Jones  
9,99 €
- 


34. In Return  
ODESZA  
9,99 €
- 

35. HVOB  
HVOB  
9,99 €
- 

36. Flume: Deluxe Edition  
Flume  
10,99 €
- 

37. Bloom  
RÜFÜS  
9,99 €
- 

38. Café del Mar 35th Anniversary  
Various Artists  
11,99 €
- 

39. Pampa, Vol. 1  
DJ Koze  
9,99 €
- 

40. Oxygen  
Fred V & Grafix  
9,99 €



```
{ '_id': ObjectId('56deffde0947000f05fc415a'),
```

```
'adamIds': [
```

```
    '1067854407',
```

1

```
    '1063750649',
```

2

```
    '1064007468',
```

3

```
    '1066300693',
```

```
    ...
```

```
    '271232254',
```

position = rank in charts

```
    '453857235',
```

```
    '377045644'
```

200

```
],
```

```
'kinds': {'album': True},
```

```
'title': 'Top Albums'
```

```
},
```

```
'discovered': 1457447797.81184,
```

```
'store-id': ,143444',
```

```
'url': 'https://itunes.../viewTop?id=27740&genreId=50'
```

```
}
```

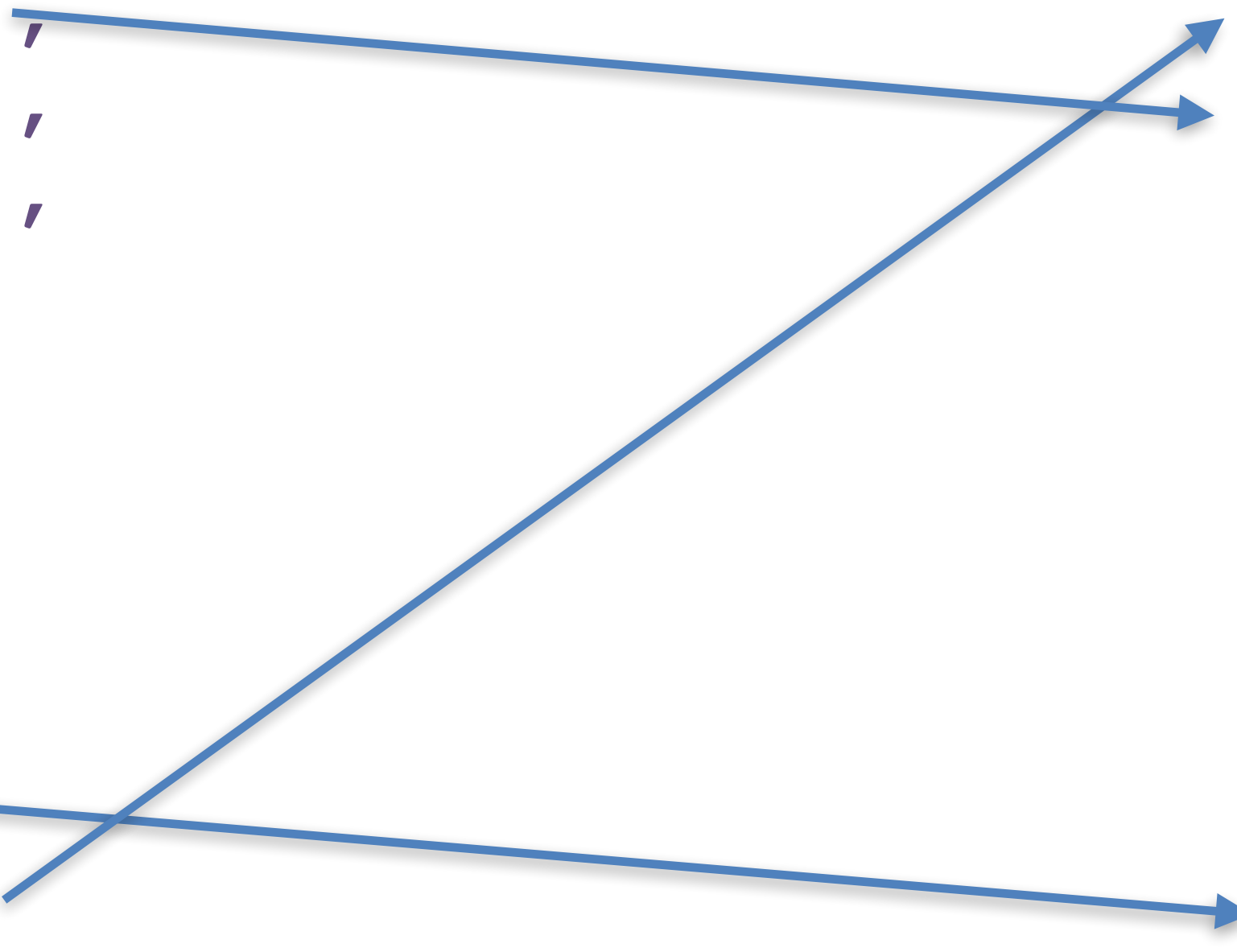
array

unix timestamp

chart id

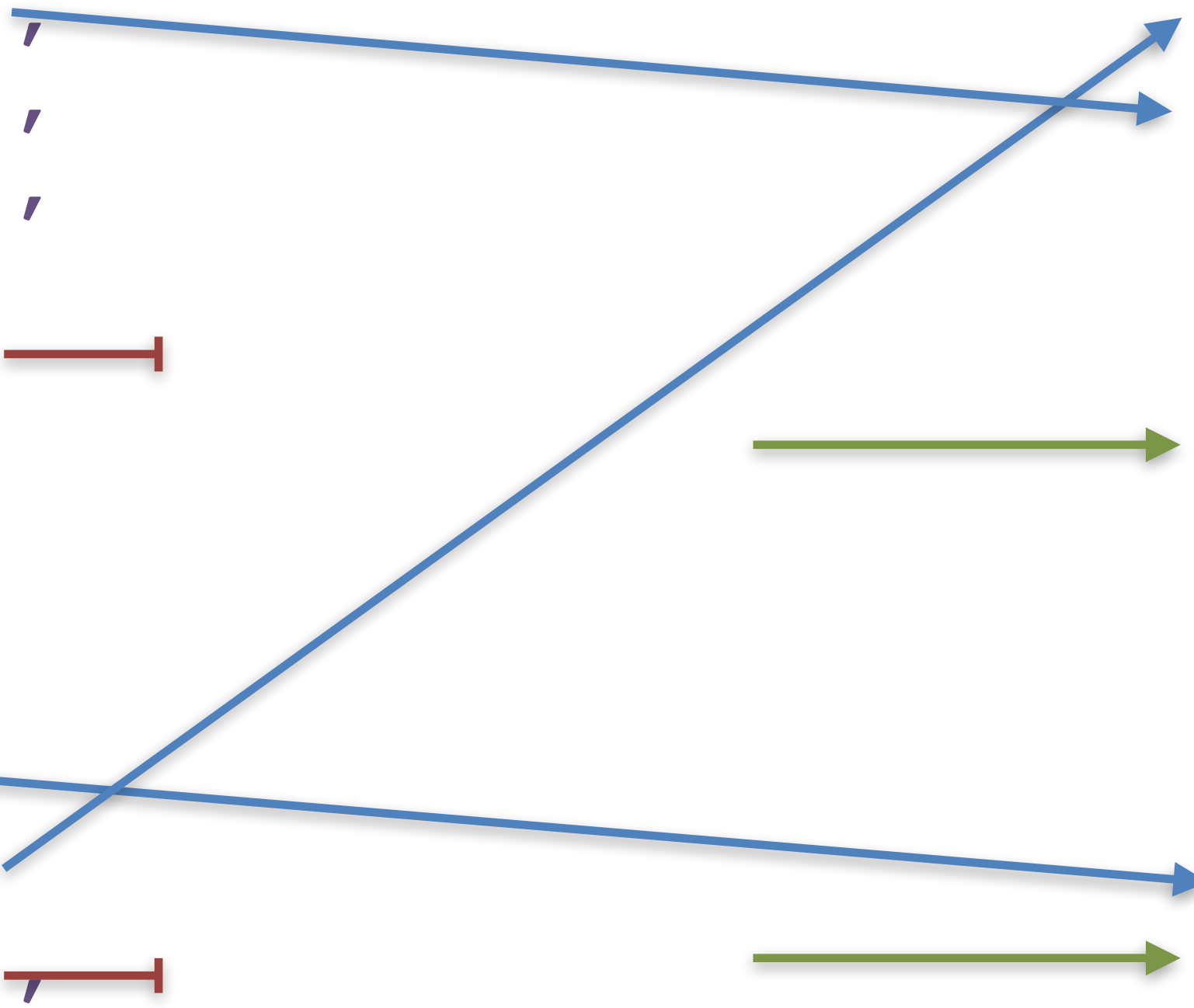
```
'adamIds': ['1067854407',  
            '1063750649',  
            '1064007468',  
            '1066300693',  
            '296867433',  
            '956751167',  
            '328069028',  
            '505586080',  
            '676328847',  
            '642644496',  
            '271232254',  
            '453857235',  
            '377045644'],  
'discovered': 1457447797
```

```
'adamIds': ['1067854407',  
            '453857235',  
            '1063750649',  
            '1066300693',  
            '296867433',  
            '328069028',  
            '292372676',  
            '505586080',  
            '676328847',  
            '642644496',  
            '956751167',  
            '271232254',  
            '544816699'],  
'discovered': 1457447836
```

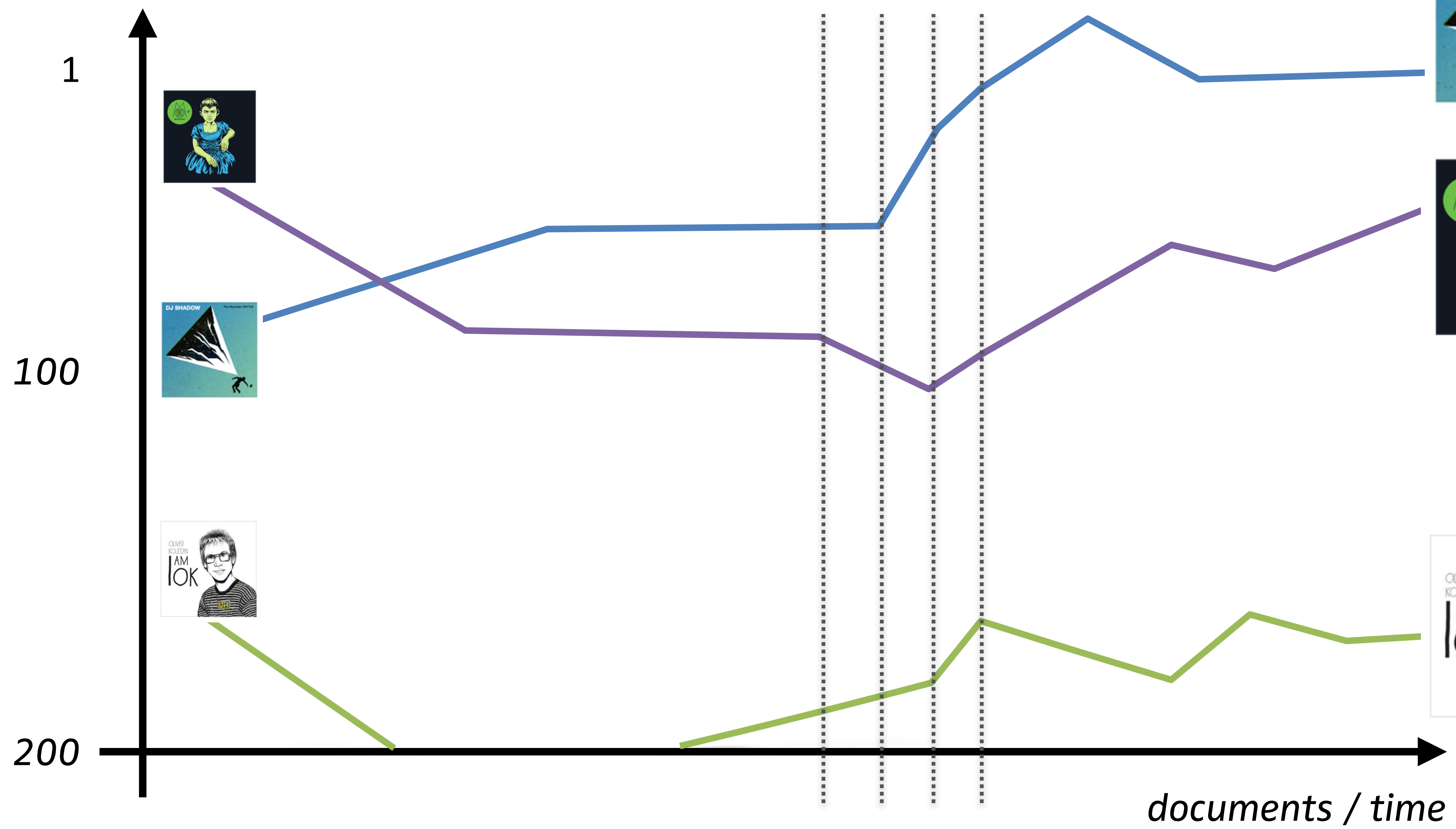


```
'adamIds': ['1067854407',  
            '1063750649',  
            '1064007468',  
            '1066300693',  
            '296867433',  
            '956751167',  
            '328069028',  
            '505586080',  
            '676328847',  
            '642644496',  
            '271232254',  
            '453857235',  
            '377045644'],  
'discovered': 1457447797
```

```
'adamIds': ['1067854407',  
            '453857235',  
            '1063750649',  
            '1066300693',  
            '296867433',  
            '328069028',  
            '292372676',  
            '505586080',  
            '676328847',  
            '642644496',  
            '956751167',  
            '271232254',  
            '544816699'],  
'discovered': 1457447836
```



rank



documents / time

```
pipeline = [  
  { "$match": {  
    "discovered": { $gte: 1457447797, $lte: 1457447836 }  
    "url": "http://the/url/is/a/identifier/the/chart/" },  
  { "$unwind": { "$adamId" } },  
  { "$group": ...  
    "$push: "$adamId" }  
]
```

```
pipeline = [ {"$match": {...}},  
{"$project": {"products": "$chart.adamIds", "discovered": "$downloadinfo.discovered"}},  
# unwind with numbering  
{"$unwind":  
  { "path": "$products", "includeArrayIndex": "arrayIndex" }},  
  
{"$project":  
  {"product": "$products",  
   # arrayIndex attribute was added by $unwind, is 0-indexed  
  "rank": {"$add": [ "$arrayIndex", 1 ]},  
  "discovered": 1,  
  # any '_id' attribute must be unique for storing, rename  
  "_id": 0, "origin_id": "$_id"}},  
{"$sort": {"origin_id": -1}},  
  
# save as new collection  
{"$out": "individual_movements" }  
]
```

```
'products': [ '1067854407',  
              '1063750649',  
              ...  
              '642644496',  
              '377045644' ],  
'discovered': 1457447797
```

```
[  {'_id': ObjectId('572c69bc8651fa448821083b'),
    'discovered': 1441110721.19208,
    'origin_id': ObjectId('55e59b260947007aef84dccb'),
    'rank': 1,
    'product': '1032438740'},
  {'_id': ObjectId('572c69bc8651fa448821083c'),
    'discovered': 1441110721.19208,
    'origin_id': ObjectId('55e59b260947007aef84dccb'),
    'rank': 2,
    'product': '976241375'}, ...
]
```

```
pipeline2 = [  
    {"$group": {"_id": "$origin_id",  
               "discovered": {"$first": "$discovered"}}},  
    {"$project": {"discovered": 1, "_id": 1}},  
    {"$out": "x_axis"}  
]
```



```
[{'_id': ObjectId('559332e6c419ab6d8b0738f9'),  
  'discovered': 1435710159.830053},  
 {'_id': ObjectId('5594ae5f09470044a56f1c61'),  
  'discovered': 1435807294.457157},  
 {'_id': ObjectId('5594bcaac419ab6d280b740e'),  
  'discovered': 1435810952.364217}]
```

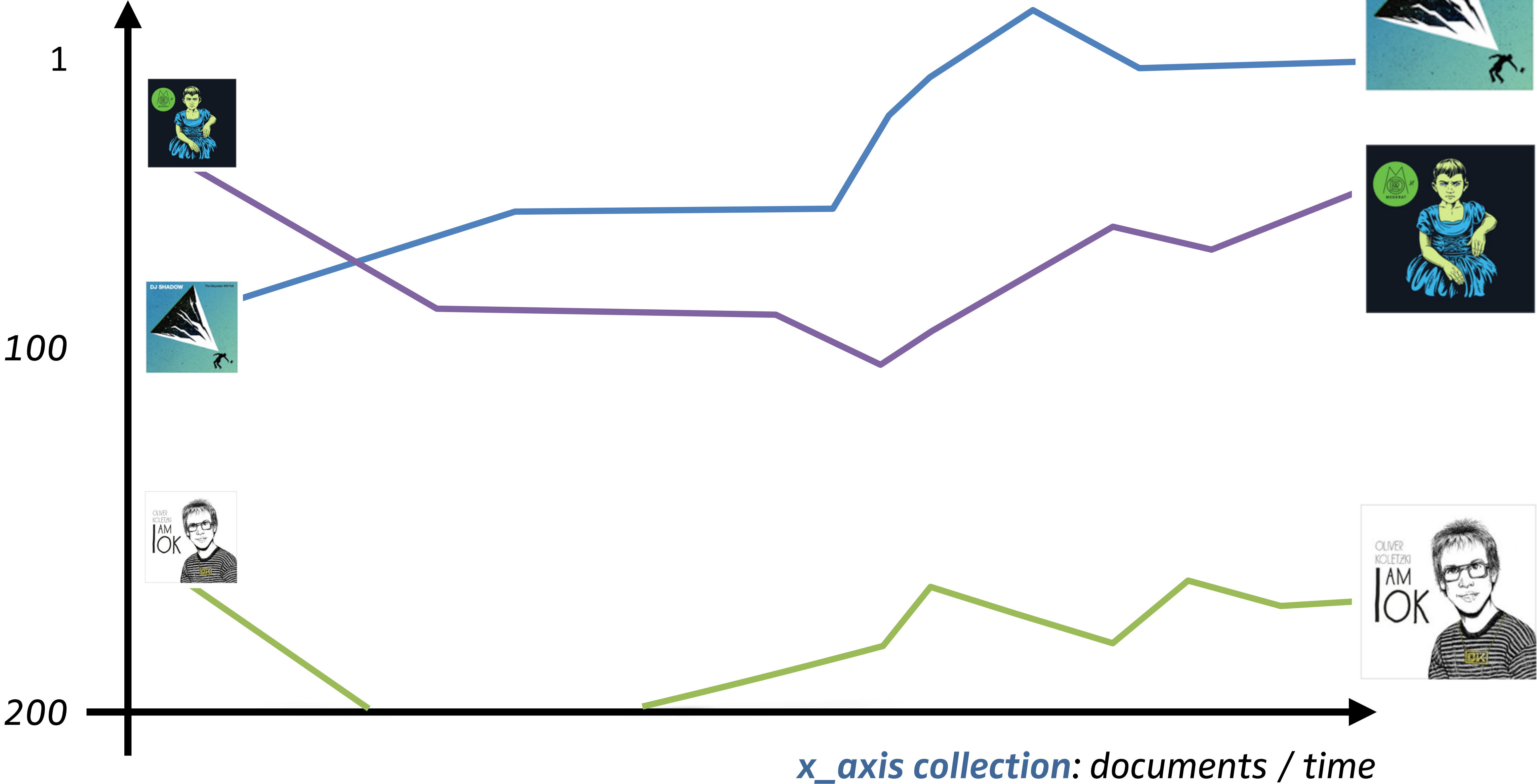
```
pipeline3 = [  
  {"$lookup": {"from": "individual_movements",  
    "localField": "_id",  
    "foreignField": "origin_id",  
    "as": "values"}},  
  
  {"$unwind": "$values"},  
  
  {"$project": {"product": "$values.product",  
    "rank": "$values.rank",  
    "discovered": 1}}  
]
```





*x\_axis collection: documents / time*

*individual\_movements: rank*

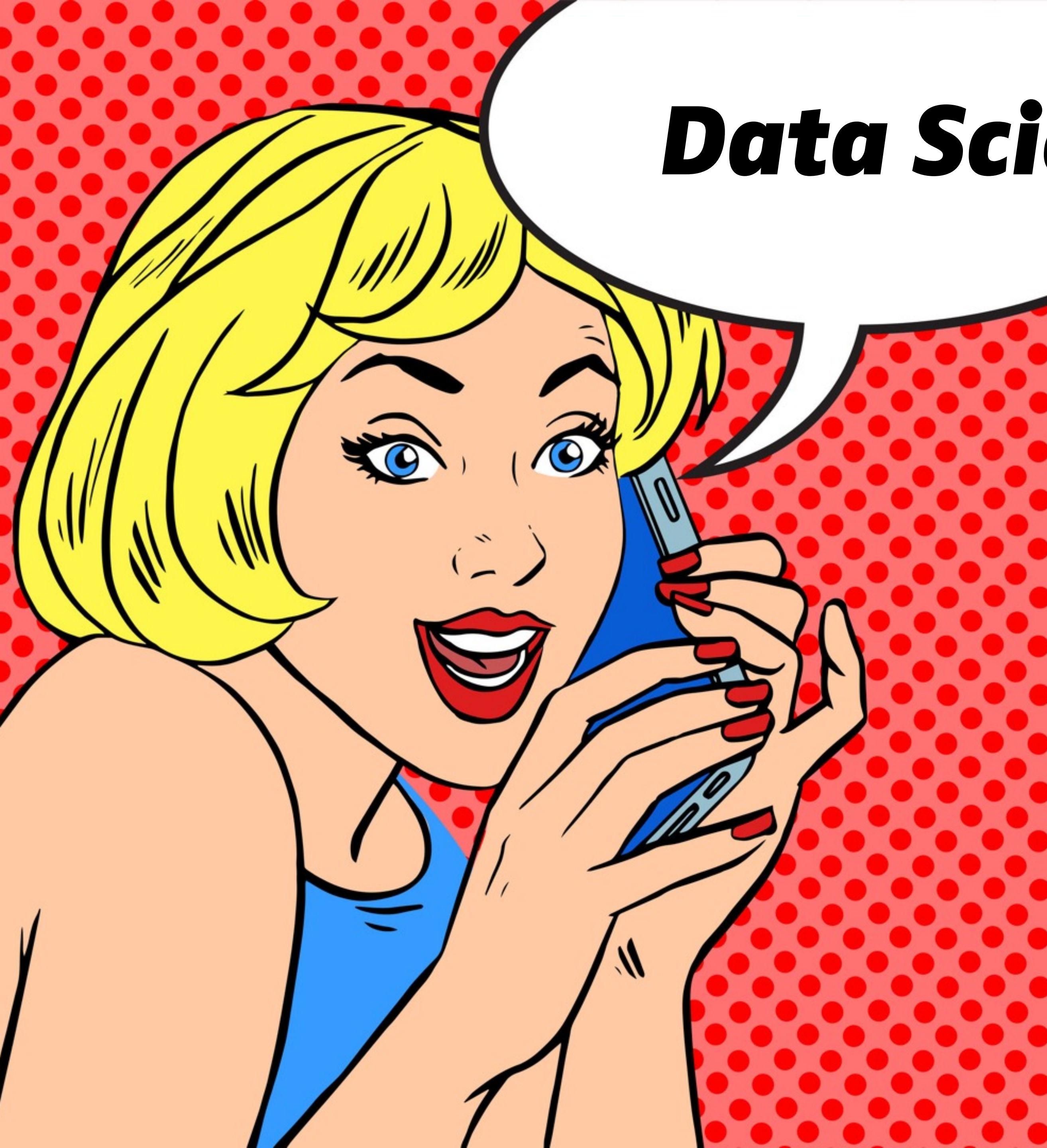


```
[{ '_id': ObjectId( '559332e6c419ab6d8b0738f9' ),  
  'discovered': 1435710159.830053,  
  'rank': 1,  
  'product': '1000697870' },  
{ '_id': ObjectId( '559332e6c419ab6d8b0738f9' ),  
  'discovered': 1435710159.830053,  
  'rank': 2,  
  'product': '986637877' },  
{ '_id': ObjectId( '559332e6c419ab6d8b0738f9' ),  
  'discovered': 1435710159.830053,  
  'rank': 3,  
  'product': '995987630' },...]
```



OK!

***Data Scientists?***







## Get Timeline / Documents from the Same Charts

```
In [ ]: match = {
    "description.genre.id": "7",
    "chart.isTracklistChart": False,
    "downloadinfo.store-id": "143443"
}
```

```
In [ ]: limit = 100
```

```
In [ ]: aggregation1 = [
    {"$match": match},
    {"$limit": limit},
    {"$project": {"products": "$chart.adamIds", "discovered": "$downloadinfo.discovered"}},
    # unwind with numbering!
    {"$unwind": {"path": "$products", "includeArrayIndex": "arrayIndex" }},
    {"$project": {"product": "$products", "rank": {"$add": ["$arrayIndex", 1]}, "discovered": 1,
        "_id": 0, "origin_id": "$_id" # exclude and rename to avoid duplicate _id in $out
    }},
    {"$sort": {"origin_id": -1}},
    {"$out": "individual_movements"},
    # Transpose, array position for object ids (sort order)
    # grouping by product, fill missing coordinates with null (or 201)
    # sort by object id -> timing, our data in recorded in chronological order
]
res = db['itunes_charts'].aggregate(aggregation1)
pprint(list(res))
print("{} datapoints recorded".format(db['individual_movements'].count()))
pprint(list(db['individual_movements'].find().limit(3)))
```

## Create X-Axis & save to temp collection

```
In [ ]: # temporal axis for all
aggregation2 = [
    {"$group": {"_id": "$origin_id",
        "discovered": {"$first": "$discovered"}}},
    {"$project": {"discovered": 1, "_id": 1}},
    {"$out": "x_axis"}
]
res = db['individual_movements'].aggregate(aggregation2)
pprint(list(res))
print("{} recordings, {}".format(db['aggregation_test_x'].count(), limit == db['x_axis'].count()))
pprint(list(db['x_axis'].find().limit(3)))
```

## Build Coordinates Data-Set



## Get Timeline / Documents from the Same Charts

```
In [ ]: match = {
    "description.genre.id": "7",
    "chart.isTracklistChart": False,
    "downloadinfo.store-id": "143443"
}
```

```
In [ ]: limit = 100
```

```
In [ ]: aggregation1 = [
    {"$match": match},
    {"$limit": limit},
    {"$project": {"products": "$chart.adamIds", "discovered": "$downloadinfo.discovered"}},
    # unwind with numbering!
    {"$unwind": {"path": "$products", "includeArrayIndex": "arrayIndex" }},
    {"$project": {"product": "$products", "rank": {"$add": ["$arrayIndex", 1]}, "discovered": 1,
        "_id": 0, "origin_id": "$_id" # exclude and rename to avoid duplicate _id in $out
    }},
    {"$sort": {"origin_id": -1}},
    {"$out": "individual_movements"},
    # Transpose, array position for object ids (sort order)
    # grouping by product, fill missing coordinates with null (or 201)
    # sort by object id -> timing, our data in recorded in chronological order
]
res = db['itunes_charts'].aggregate(aggregation1)
pprint(list(res))
print("{} datapoints recorded".format(db['individual_movements'].count()))
pprint(list(db['individual_movements'].find().limit(3)))
```

## Create X-Axis & save to temp collection

```
In [ ]: # temporal axis for all
aggregation2 = [
    {"$group": {"_id": "$origin_id",
        "discovered": {"$first": "$discovered"}}},
    {"$project": {"discovered": 1, "_id": 1}},
    {"$out": "x_axis"}
]
res = db['individual_movements'].aggregate(aggregation2)
pprint(list(res))
print("{} recordings, {}".format(db['aggregation_test_x'].count(), limit == db['x_axis'].count()))
pprint(list(db['x_axis'].find().limit(3)))
```

## Build Coordinates Data-Set



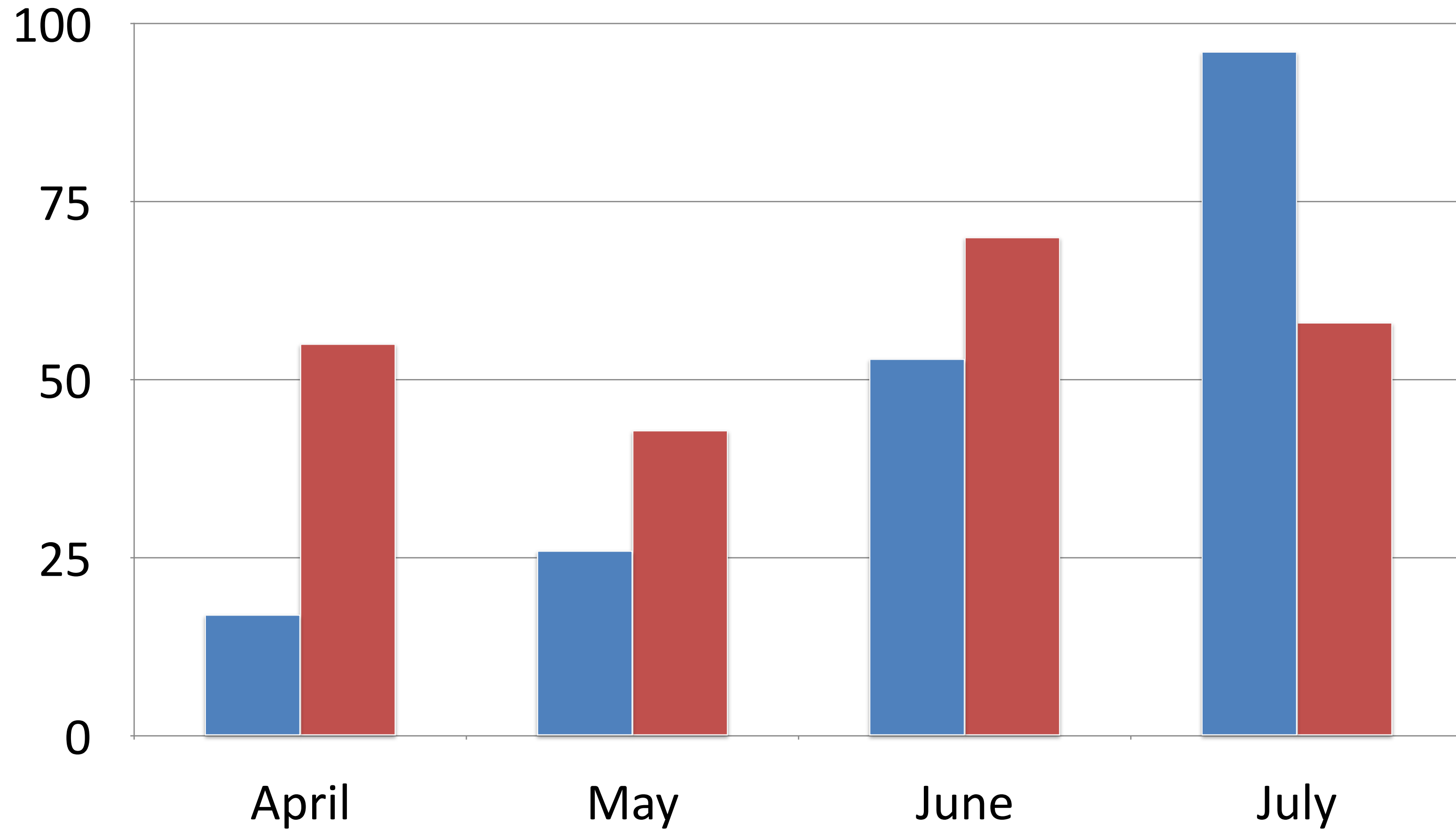
# ***Data Scientists!***

- Grant access with the built in **role management**
- Data scientists can analyse the data with typical tools as Pandas, R, etc...
- easy as a cake with VIEWS coming in 3.4 :-)

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Auftrags-ID	Bestelldatum	Versanddatum	Versandmodus	Kunden-ID	Kundenname	Segment	Stadt	Bundesland	Land	Region	Produkt-ID	Kategorie	Unterkategorie	Produktname	Umsatz	Me
ES-2014-1311038	07.02.14	11.02.14	Standardklasse	AN-10045	Alexander Nadel	Großunternehmen	Leeds	England	Vereinigtes Königreich	Nord	BÜR-AU-1000290	Bürobedarf	Aufbewahrung	Fellowes Ordner, Blau	79,2	
ES-2014-1311038	07.02.14	11.02.14	Standardklasse	AN-10045	Alexander Nadel	Großunternehmen	Leeds	England	Vereinigtes Königreich	Nord	TEC-CO-1000452	Technik	Computerzubehör	SanDisk Tastenfeld, Haltbar	388,92	
ES-2014-1311038	07.02.14	11.02.14	Standardklasse	AN-10045	Alexander Nadel	Großunternehmen	Leeds	England	Vereinigtes Königreich	Nord	BÜR-ET-1000400	Bürobedarf	Etiketten	Avery Asservaten-Etiketten, Haltbar	35,19	
ES-2014-1311038	07.02.14	11.02.14	Standardklasse	AN-10045	Alexander Nadel	Großunternehmen	Leeds	England	Vereinigtes Königreich	Nord	BÜR-AU-1000195	Bürobedarf	Aufbewahrung	Fellowes Ordner, Metall	50,94	
ES-2014-1311038	07.02.14	11.02.14	Standardklasse	AN-10045	Alexander Nadel	Großunternehmen	Leeds	England	Vereinigtes Königreich	Nord	TEC-CO-1000278	Technik	Computerzubehör	Memorex Speicherkarte, Haltbar	307,44	
ES-2014-1311038	07.02.14	11.02.14	Standardklasse	AN-10045	Alexander Nadel	Großunternehmen	Leeds	England	Vereinigtes Königreich	Nord	BÜR-AU-1000485	Bürobedarf	Aufbewahrung	Rogers Regale, Metall	122,4	
ES-2014-1311038	07.02.14	11.02.14	Standardklasse	AN-10045	Alexander Nadel	Großunternehmen	Leeds	England	Vereinigtes Königreich	Nord	TEC-TE-1000317	Technik	Telefone	Apple Signalverstärker, Fullsize	413,82	
ES-2012-5113958	01.08.12	06.08.12	Zweite Klasse	JS-13840	Jessika Schmidt	Großunternehmen	West Bromwich	England	Vereinigtes Königreich	Nord	TEC-KO-1000201	Technik	Kopierer	Canon Tischkopierer, Zweierset	428,22	
ES-2012-5113958	01.08.12	06.08.12	Zweite Klasse	JS-13840	Jessika Schmidt	Großunternehmen	West Bromwich	England	Vereinigtes Königreich	Nord	BÜR-GE-1000144	Bürobedarf	Geräte	Hoover Herd, Rot	3979,29	
ES-2012-5113958	01.08.12	06.08.12	Zweite Klasse	JS-13840	Jessika Schmidt	Großunternehmen	West Bromwich	England	Vereinigtes Königreich	Nord	BÜR-BE-1000003	Bürobedarf	Klemmen	Accos Pinnadeln, Großpackung	43,56	
ES-2012-2510515	19.06.12	19.06.12	Am gleichen Tag	MG-17155	Mario Gersten	Verbraucher	Le Bouscat	Aquitaine-Limousin	Frankreich	Zentral	BÜR-ET-1000063	Bürobedarf	Etiketten	Hon Farbcodierte Etiketten, Rot	25,26	
ES-2012-2510515	19.06.12	19.06.12	Am gleichen Tag	MG-17155	Mario Gersten	Verbraucher	Le Bouscat	Aquitaine-Limousin	Frankreich	Zentral	BÜR-GE-1000335	Bürobedarf	Geräte	Hamilton Beach Herd, Silber	2443,905	
ES-2012-2510515	19.06.12	19.06.12	Am gleichen Tag	MG-17155	Mario Gersten	Verbraucher	Le Bouscat	Aquitaine-Limousin	Frankreich	Zentral	BÜR-KU-1000246	Bürobedarf	Künstlerbedarf	Binney & Smith Stifte, Großpackung	12,21	
ES-2012-2510515	19.06.12	19.06.12	Am gleichen Tag	MG-17155	Mario Gersten	Verbraucher	Le Bouscat	Aquitaine-Limousin	Frankreich	Zentral	TEC-TE-1000126	Technik	Telefone	Samsung Ladegerät, Fullsize	2167,296	
ES-2012-2510515	19.06.12	19.06.12	Am gleichen Tag	MG-17155	Mario Gersten	Verbraucher	Le Bouscat	Aquitaine-Limousin	Frankreich	Zentral	BÜR-AU-1000444	Bürobedarf	Aufbewahrung	Rogers Ordner, Blau	138,105	
ES-2013-4380115	15.02.13	19.02.13	Standardklasse	MM-16540	Maik Müller	Verbraucher	Straßburg	Alsace-Champagne	Frankreich	Zentral	BÜR-AU-1000107	Bürobedarf	Aufbewahrung	Smead Tablett, Metall	128,385	
ES-2013-4380115	15.02.13	19.02.13	Standardklasse	MM-16540	Maik Müller	Verbraucher	Straßburg	Alsace-Champagne	Frankreich	Zentral	TEC-CO-1000466	Technik	Computerzubehör	Belkin Speicherkarte, Haltbar	690,12	
ES-2013-4380115	15.02.13	19.02.13	Standardklasse	MM-16540	Maik Müller	Verbraucher	Straßburg	Alsace-Champagne	Frankreich	Zentral	BÜR-ET-1000134	Bürobedarf	Etiketten	Hon Entfernbare Etiketten, Einstellbar	8,16	
ES-2014-2573330	03.11.14	08.11.14	Standardklasse	CJ-11170	Christian Jung	Kleinunternehmen	Unna	Nordrhein-Westfalen	Deutschland	Zentral	TEC-GE-1000336	Technik	Maschinen	Okidata Telefon, Haltbar	313,092	
ES-2014-2573330	03.11.14	08.11.14	Standardklasse	CJ-11170	Christian Jung	Kleinunternehmen	Unna	Nordrhein-Westfalen	Deutschland	Zentral	BÜR-AU-1000065	Bürobedarf	Aufbewahrung	Eldon Aktenwagen, Blau	575,505	
ES-2014-1872792	14.08.14	16.08.14	Erste Klasse	CV-11275	Christian Vogler	Großunternehmen	Thiais	Ile-de-France	Frankreich	Zentral	BÜR-OR-1000255	Bürobedarf	Ordner	Acco Lochverstärkungen, Recycelt	14,04	
ES-2014-3412387	18.06.14	24.06.14	Standardklasse	FV-12880	Frank Vogel	Großunternehmen	Kassel	Hessen	Deutschland	Zentral	MÖB-BÜ-100008	Möbel	Bücherregale	Safco Bücherregal, Traditionell	268,164	
IT-2014-3245851	02.05.14	05.05.14	Zweite Klasse	MB-16795	Marco Boehm	Kleinunternehmen	Mailand	Italien	Süd	BÜR-AU-1000485	Bürobedarf	Aufbewahrung	Rogers Regale, Metall	36,72		
IT-2014-3245851	02.05.14	05.05.14	Zweite Klasse	MB-16795	Marco Boehm	Kleinunternehmen	Mailand	Italien	Süd	TEC-CO-1000310	Technik	Computerzubehör	Memorex Maus, Preiswert	28,17		
IT-2013-1215824	12.08.13	16.08.13	Standardklasse	SK-19780	Sophie Krüger	Verbraucher	Brandenburg	Brandenburg	Deutschland	Zentral	BÜR-AU-1000485	Bürobedarf	Aufbewahrung	Rogers Regale, Fullsize	279,72	
IT-2013-1215824	12.08.13	16.08.13	Standardklasse	SK-19780	Sophie Krüger	Verbraucher	Brandenburg	Brandenburg	Deutschland	Zentral	BÜR-KU-1000417	Bürobedarf	Künstlerbedarf	BIC Stifte, Metall	108,78	
ES-2014-2557751	28.02.14	04.03.14	Standardklasse	TB-20920	Tobias Bach	Verbraucher	Heilbronn	Baden-Württemberg	Deutschland	Zentral	TEC-GE-1000187	Technik	Geräte	Panasonic Rechner, Rot	298,809	
ES-2014-2557751	28.02.14	04.03.14	Standardklasse	TB-20920	Tobias Bach	Verbraucher	Heilbronn	Baden-Württemberg	Deutschland	Zentral	BÜR-AU-1000175	Bürobedarf	Aufbewahrung	Rogers Karton, Industriell	110,025	
ES-2014-2557751	28.02.14	04.03.14	Standardklasse	TB-20920	Tobias Bach	Verbraucher	Heilbronn	Baden-Württemberg	Deutschland	Zentral	BÜR-BE-1000414	Bürobedarf	Klemmen	Stockwell Gummibänder, Verschiedene	41,94	
ES-2014-4134792	11.09.14	15.09.14	Standardklasse	AK-10750	Antje Koenig	Verbraucher	Herten	Nordrhein-Westfalen	Deutschland	Zentral	TEC-CO-1000273	Technik	Computerzubehör	Belkin Maus, Recycelt	192,6	
ES-2014-4134792	11.09.14	15.09.14	Standardklasse	AK-10750	Antje Koenig	Verbraucher	Herten	Nordrhein-Westfalen	Deutschland	Zentral	TEC-TE-1000181	Technik	Telefone	Samsung Audio-Dock, Fullsize	336,24	
ES-2011-2801336	01.06.11	05.06.11	Standardklasse	IK-13270	Ines Kaiser	Kleinunternehmen	Barcelona	Katalonien	Spanien	Süd	BÜR-OR-1000156	Bürobedarf	Ordner	Cardinal Bindemaschine, Haltbar	102,54	
ES-2013-5191940	14.08.13	19.08.13	Standardklasse	MB-16795	Marco Boehm	Kleinunternehmen	Swindon	England	Vereinigtes Königreich	Nord	BÜR-PA-1000238	Bürobedarf	Papier	Green Bar Memo-Blätter, 12er-Pack	63,27	
ES-2013-5191940	14.08.13	19.08.13	Standardklasse	MB-16795	Marco Boehm	Kleinunternehmen	Swindon	England	Vereinigtes Königreich	Nord	BÜR-PA-1000157	Bürobedarf	Papier	Eaton Packung, Mehrfarbig	55,92	
ES-2013-4020783	25.11.13	27.11.13	Zweite Klasse	RK-18775	Ralf Kruger	Kleinunternehmen	Puertollano	Kastilien-La Mancha	Spanien	Süd	BÜR-KU-1000421	Bürobedarf	Künstlerbedarf	Stanley Marker, Verschiedene Größen	76,14	
ES-2013-3284813	31.05.13	05.06.13	Standardklasse	SR-20020	Stefanie Reiniger	Verbraucher	Chaville	Ile-de-France	Frankreich	Zentral	BÜR-KU-1000106	Bürobedarf	Künstlerbedarf	BIC Skizzenblock, Verschiedene Größen	155,52	
ES-2013-3284813	31.05.13	05.06.13	Standardklasse	SR-20020	Stefanie Reiniger	Verbraucher	Chaville	Ile-de-France	Frankreich	Zentral	BÜR-AU-1000027	Bürobedarf	Aufbewahrung	Eldon Karton, Blau	46,17	
ES-2013-3284813	31.05.13	05.06.13	Standardklasse	SR-20020	Stefanie Reiniger	Verbraucher	Chaville	Ile-de-France	Frankreich	Zentral	BÜR-ET-1000239	Bürobedarf	Etiketten	Novimex Farbcodierte Etiketten, Weiß	37,26	
IT-2014-5016792	10.01.14	12.01.14	Zweite Klasse	UB-21280	Ulrich Bergmann	Verbraucher	Sonneberg	Thüringen	Deutschland	Zentral	BÜR-PA-1000238	Bürobedarf	Papier	Eaton Memo-Blätter, Mehrfarbig	90,9	
IT-2014-5016792	10.01.14	12.01.14	Zweite Klasse	UB-21280	Ulrich Bergmann	Verbraucher	Sonneberg	Thüringen	Deutschland	Zentral	BÜR-KU-1000262	Bürobedarf	Künstlerbedarf	Binney & Smith Stifte, Verschiedene Größen	75,6	
IT-2013-2373016	08.11.13	11.11.13	Erste Klasse	DP-11830	David Papst	Großunternehmen	Den Haag	Südholland	Niederlande	Zentral	BÜR-ZU-1000231	Bürobedarf	Zubehör	Stiletto Brieföffner, Industriell	15	
IT-2013-2373016	08.11.13	11.11.13	Erste Klasse	DP-11830	David Papst	Großunternehmen	Den Haag	Südholland	Niederlande	Zentral	BÜR-OR-1000255	Bürobedarf	Ordner	Acco Lochverstärkungen, Recycelt	14,04	
IT-2013-2373016	08.11.13	11.11.13	Erste Klasse	DP-11830	David Papst	Großunternehmen	Den Haag	Südholland	Niederlande	Zentral	BÜR-KU-1000467	Bürobedarf	Künstlerbedarf	Stanley Stifte, Verschiedene Größen	33,075	
ES-2011-3186072	27.09.11	01.10.11	Standardklasse	DM-12265	Dirk Mauer	Verbraucher	Littlehampton	England	Vereinigtes Königreich	Nord	TEC-KO-1000400	Technik	Kopierer	Brother Drahtlos-Fax, Digital	1892,25	
ES-2011-3186072	27.09.11	01.10.11	Standardklasse	DM-12265	Dirk Mauer	Verbraucher	Littlehampton	England	Vereinigtes Königreich	Nord	TEC-CO-1000305	Technik	Computerzubehör	Logitech Flash-Laufwerk, Preiswert	122,04	
ES-2011-3186072	27.09.11	01.10.11	Standardklasse	DM-12265	Dirk Mauer	Verbraucher	Littlehampton	England	Vereinigtes Königreich	Nord	TEC-KO-1000137	Technik	Kopierer	Brother Drahtlos-Fax, Farbe	1140,75	
ES-2011-5235241	26.08.11	27.08.11	Am gleichen Tag	TG-21175	Torsten Gerber	Großunternehmen	Taverny	Ile-de-France	Frankreich	Zentral	BÜR-ZU-1000411	Bürobedarf	Zubehör	Stiletto Beschneidemaschine, Stahl	84,3	
ES-2011-5235241	26.08.11	27.08.11	Am gleichen Tag	TG-21175	Torsten Gerber	Großunternehmen	Taverny	Ile-de-France	Frankreich	Zentral	TEC-KO-1000245	Technik	Kopierer	Canon Faxgerät, Rot	538,203	
ES-2011-5235241	26.08.11	27.08.11	Am gleichen Tag	TG-21175	Torsten Gerber	Großunternehmen	Taverny	Ile-de-France	Frankreich	Zentral	BÜR-KU-1000091	Bürobedarf	Künstlerbedarf	Binney & Smith Leinwand, Metall	157,86	
ES-2014-5640080	16.04.14	20.04.14	Standardklasse	ME-17455	Martin Eberhardt	Verbraucher	Mainz	Rheinland-Pfalz	Deutschland	Zentral	BÜR-ZU-1000165	Bürobedarf	Zubehör	Fiskars Cuttermesser, Blau	106,56	
IT-2014-2321796	02.04.14	07.04.14	Standardklasse	KE-15100	Kludia Ebersbach	Verbraucher	Nacka	Stockholms	Schweden	Nord	BÜR-AU-1000332	Bürobedarf	Aufbewahrung	Smead Regale, Metall	72	
ES-2014-4777800	26.11.14	29.11.14	Erste Klasse	SF-19330	Sandra Fürst	Verbraucher	Manchester	England	Vereinigtes Königreich	Nord	BÜR-KU-1000122	Bürobedarf	Künstlerbedarf	Boston Leinwand, Verschiedene Größen	227,28	
ES-2014-3937023	07.11.14	11.11.14	Standardklasse	LI-15715	Lea Lehrer	Verbraucher	Viroflay	Ile-de-France	Frankreich	Zentral	BÜR-KU-1000346	Bürobedarf	Künstlerbedarf	Stanley Marker, Metall	46,92	

Data

# ***Visualization!***



***Analysts?***



# BI Connector

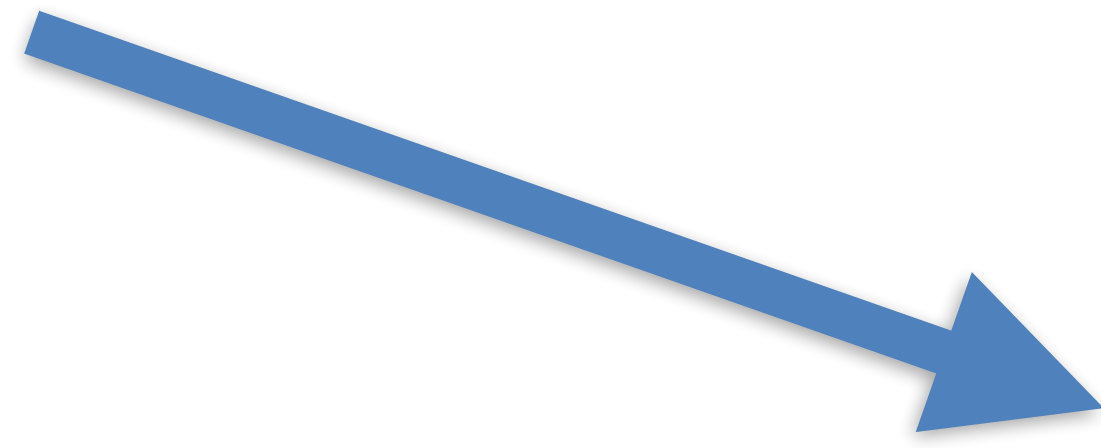






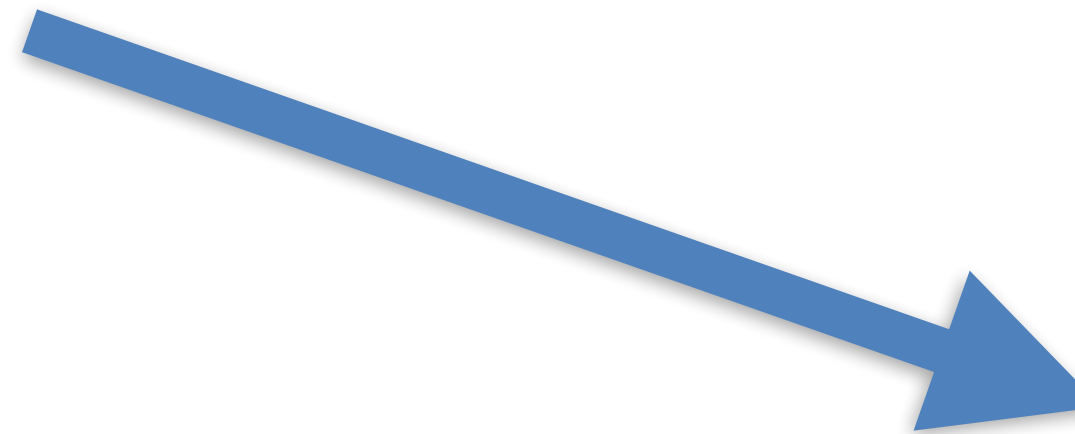
```
{ '_id': ObjectId( '559332e6c419ab6d8b0738f9' ),  
  'rank': 0,  
  'product': '1000697870':  
  'abc': [ "a", "b", "c" ]  
}
```

```
{ '_id': ObjectId( '559332e6c419ab6d8b0738f9' ),  
  'rank': 0,  
  'product': '1000697870':  
  'abc': [ "a", "b", "c" ]  
}
```



```
schema:  
- db: mydatabase  
tables:  
- table: my_mongodb_collection  
collection: my_mongodb_collection  
pipeline: []  
columns:  
- Name: _id  
  MongoType: bson.ObjectId  
  SqlName: _id  
  SqlType: varchar  
- Name: rank  
  MongoType: int  
  SqlName: rank  
  SqlType: numeric  
- Name: product  
  MongoType: string  
  SqlName: product  
  SqlType: varchar  
...
```

```
{ '_id': ObjectId( '559332e6c419ab6d8b0738f9' ),  
  'rank': 0,  
  'product': '1000697870':  
  'abc': [ "a", "b", "c" ]  
}
```



```
SqlType: varchar  
- Name: rank  
MongoType: int  
SqlName: rank  
SqlType: numeric  
- Name: product  
MongoType: string  
SqlName: product  
SqlType: varchar
```

...

```
- table: my_mongodb_collection_abc  
collection: my_mongodb_collection  
pipeline:  
- $unwind:  
  includeArrayIndex: abc  
  path: $abc  
columns:  
- Name: abc  
MongoType: string  
SqlName: abc
```

...

```
mongobiuser create user mongodb://localhost:27017/myDB  
# create a user account
```


```
mongodrdl --host localhost -d myDB -o schema.drdl  
# create the schema
```

```
mongobischema import user schema.drdl  
# load the schema
```

# BI Connector







# CITY of BOSTON

## Employee Demographics

Last Updated: 23. Mai 2016

### Gender

**Overall** Male 46,4% Female 53,6%


#### By Annual Rate

Annual Rate	Male	Female
0 - 20k	26,9%	73,1%
20 - 40k	37,1%	62,9%
40 - 60k	47,5%	52,5%
60 - 80k	54,6%	45,4%
80 - 100k	53,6%	46,4%
> 100k	46,7%	53,3%

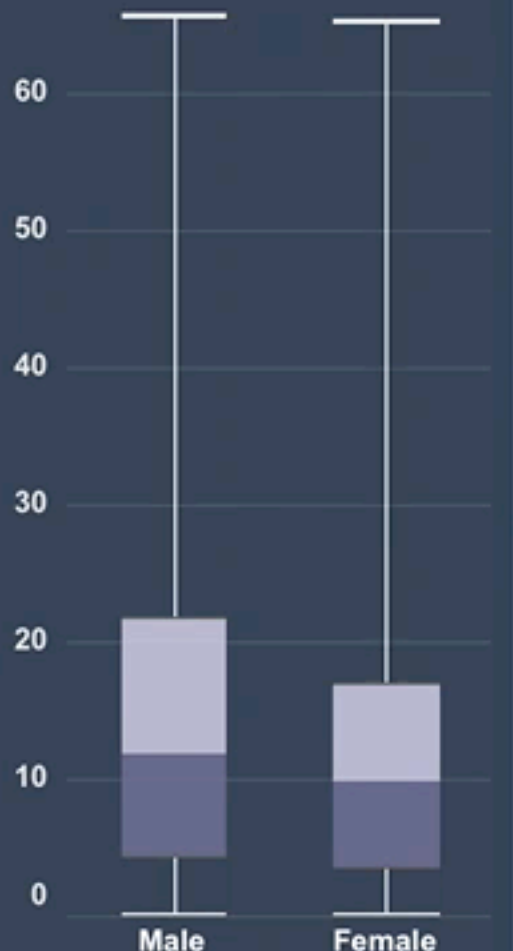
#### By Ethnic Group

Ethnic Group	Male	Female
American Ind	41,0%	59,0%
Asian	35,0%	65,0%
Black	43,6%	56,4%
Hispanic	37,6%	62,4%
Pacific	41,2%	58,8%
White	50,9%	49,1%
Unspecified	31,4%	68,6%

### Annual Rate



### Tenure - Years



Full-Time 72,5%  
Part-Time 5,2%  
Temporary 22,3%

*Click to filter. Click again to release filter.*

#### Filter by Service Date

12.12.1950 — 23.05.2016

BPS - Boston Public Schools  
 COB - City of Boston

#### Search by Department Name

#### Filter by Department

- (Alle)
- BPS - Achievement Gap
- BPS - Alighieri Montessori School
- BPS - Asst Superintendent-Network A
- BPS - Asst Superintendent-Network B
- BPS - Asst Superintendent-Network C
- BPS - Asst Superintendent-Network D
- BPS - Asst Superintendent-Network E
- BPS - Asst Superintendent-Network F
- BPS - Asst Superintendent-Network G
- BPS - Baldwin ELC
- BPS - BPS Adams Elementary
- BPS - BPS Adult Education
- BPS - BPS Alternative Education
- BPS - BPS Alternative Education HS
- BPS - BPS Another Course To Colleg
- BPS - BPS Athletics
- BPS - BPS Bates Elementary
- BPS - BPS Beethoven Elementary
- BPS - BPS Blackstone Elementary
- BPS - BPS Boston Arts Academy
- BPS - BPS Boston Comm Leadership Ac
- BPS - BPS Boston Evening Academy
- BPS - BPS Boston International HS
- BPS - BPS Boston Latin

*Please note: The information displayed on this page was collected pursuant to and consistent with Federal EEOC rules and requirements.*


### Ethnicity

**Overall** White 54,6% Black 27,9% Hispan... 11,9% Asian 4,0% Unspe... 1,3% Americ... 0,2% Pacific 0,1%

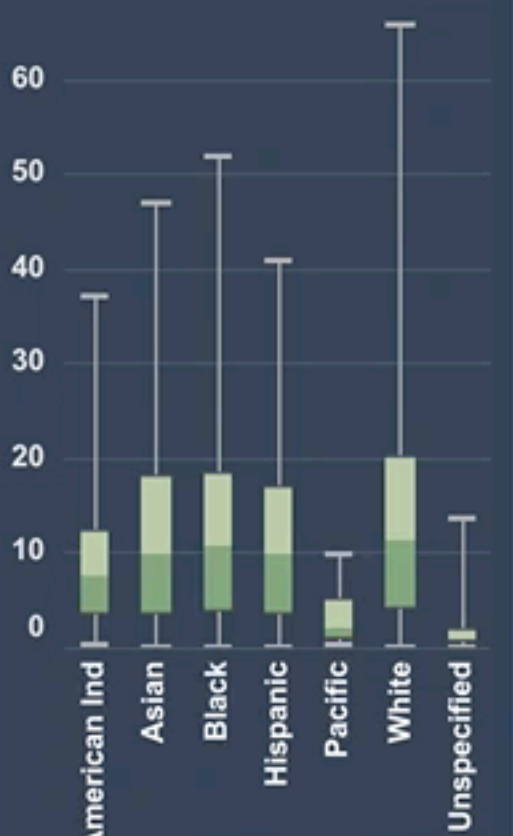
#### By Annual Rate

Annual Rate	White	Black	Hispan...	Asian	Unspe...	Americ...	Pacific
0 - 20k	34,0%	41,4%	16,2%	3,4%	4,5%	0,3%	0,2%
20 - 40k	36,3%	38,4%	20,5%	2,9%	1,5%	0,2%	0,2%
40 - 60k	50,4%	31,2%	12,7%	4,3%	1,1%	0,2%	0,1%
60 - 80k	63,0%	21,5%	9,5%	4,7%	1,1%	0,1%	0,1%
80 - 100k	62,7%	23,6%	9,4%	3,8%	0,1%	0,3%	
> 100k	68,5%	18,6%	7,8%	4,5%	0,3%	0,2%	0,0%

### Annual Rate



### Tenure - Years



# CITY of BOSTON

## Employee Demographics

Last Updated: 23. Mai 2016

### Gender

**Overall** Male 46,4% Female 53,6%

#### By Annual Rate

Annual Rate	Male	Female
0 - 20k	26,9%	73,1%
20 - 40k	37,1%	62,9%
40 - 60k	47,5%	52,5%
60 - 80k	54,6%	45,4%
80 - 100k	53,6%	46,4%
> 100k	46,7%	53,3%

#### By Ethnic Group

Ethnic Group	Male	Female
American Ind	41,0%	59,0%
Asian	35,0%	65,0%
Black	43,6%	56,4%
Hispanic	37,6%	62,4%
Pacific	41,2%	58,8%
White	50,9%	49,1%
Unspecified	31,4%	68,6%

### Annual Rate

Box plot showing Annual Rate (0K to 250K) for Male and Female. Male rates are generally higher than female rates.

### Tenure - Years

Box plot showing Tenure in years (0 to 60) for Male and Female. Male tenure is generally higher than female tenure.

Full-Time 72,5%  
Part-Time 5,2%  
Temporary 22,3%

*Click to filter. Click again to release filter.*

#### Filter by Service Date

12.12.1950 — 23.05.2016

BPS - Boston Public Schools  
 COB - City of Boston

#### Search by Department Name

#### Filter by Department

- (Alle)
- BPS - Achievement Gap
- BPS - Alighieri Montessori School
- BPS - Asst Superintendent-Network A
- BPS - Asst Superintendent-Network B
- BPS - Asst Superintendent-Network C
- BPS - Asst Superintendent-Network D
- BPS - Asst Superintendent-Network E
- BPS - Asst Superintendent-Network F
- BPS - Asst Superintendent-Network G
- BPS - Baldwin ELC
- BPS - BPS Adams Elementary
- BPS - BPS Adult Education
- BPS - BPS Alternative Education
- BPS - BPS Alternative Education HS
- BPS - BPS Another Course To Colleg
- BPS - BPS Athletics
- BPS - BPS Bates Elementary
- BPS - BPS Beethoven Elementary
- BPS - BPS Blackstone Elementary
- BPS - BPS Boston Arts Academy
- BPS - BPS Boston Comm Leadership Ac
- BPS - BPS Boston Evening Academy
- BPS - BPS Boston International HS
- BPS - BPS Boston Latin

*Please note: The information displayed on this page was collected pursuant to and consistent with Federal EEOC rules and requirements.*

### Ethnicity

**Overall** White 54,6% Black 27,9% Hispan... 11,9% Asian 4,0% Unspe... 1,3% Americ... 0,2% Pacific 0,1%

#### By Annual Rate

Annual Rate	White	Black	Hispan...	Asian	Unspe...	Americ...	Pacific
0 - 20k	34,0%	41,4%	16,2%	3,4%	4,5%	0,3%	0,2%
20 - 40k	36,3%	38,4%	20,5%	2,9%	1,5%	0,2%	0,2%
40 - 60k	50,4%	31,2%	12,7%	4,3%	1,1%	0,2%	0,1%
60 - 80k	63,0%	21,5%	9,5%	4,7%	1,1%	0,1%	0,1%
80 - 100k	62,7%	23,6%	9,4%	3,8%	0,1%	0,3%	
> 100k	68,5%	18,6%	7,8%	4,5%	0,3%	0,2%	0,0%

### Annual Rate

Box plot showing Annual Rate (0K to 250K) for various ethnic groups. White employees generally have the highest annual rates.

### Tenure - Years

Box plot showing Tenure in years (0 to 60) for various ethnic groups. White employees generally have the highest tenure.



***Analysts!***

## BI Connector

- access mongoDB from BI tools
- mock-up of RDBS
- BI user accounts







**Alexander C. S. Hendorf**

 @hendorf