MIDTERM ASSIGNMENT

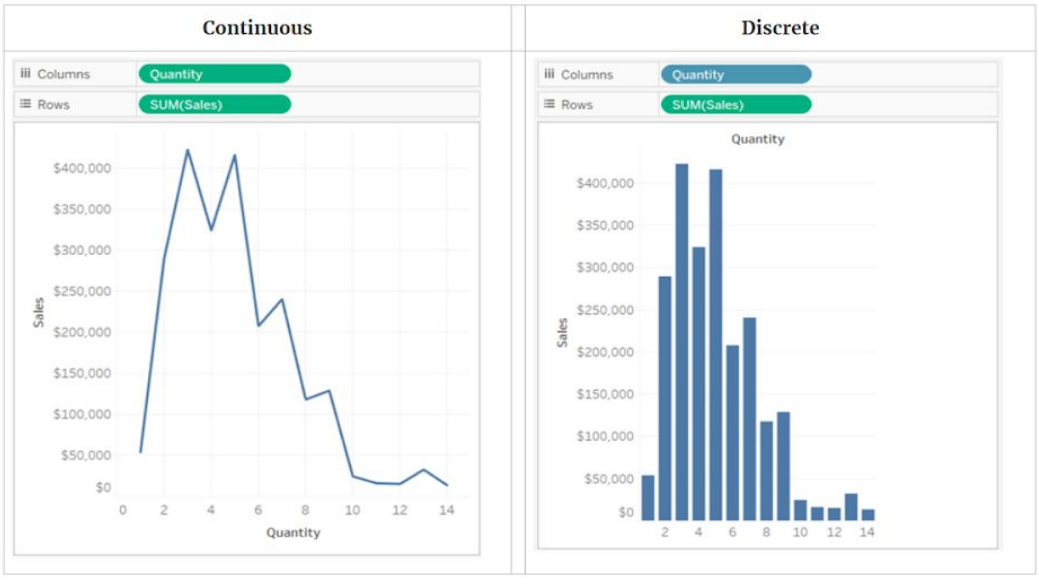
PART I: Fundamental Concepts

1. Once after connecting to a data source, Tableau will classify each attribute either as a dimension or a measure in the Data pane. It depends on the data type of the attribute such as string, date, int.
   - **Dimensions**: Dimensions are considered as independent variables. They contain qualitative and categorical information like IDs, Names, Dates, Product categories, Geographical data. These fields cannot be measured or aggregated. These dimensions are used in the visualizations to segment, categorize and show the details within the data. They will affect the level of detail in the view.
   - **Measures**: Measure is a field that is a dependent variable. It is dependent on the context that comes in the form of being broken down by dimensions.They contain quantitative values (numeric) that we can measure. They can be aggregated as sum, average, median, count, and count distinct. In tableau, by default an aggregation is applied to a measure when it is dragged into the view.

   Generally, a measure is a number, and the dimension is what we slice and dice the number by. When a field is dragged into the view, tableau represents them differently based on whether the field is continuous (green), or discrete (blue).

   - Both the green dimensions YEAR(Order Date) and measures SUM(Profit) are continuous. Continuous fields will add axes to the view and the values are considered as an infinite range.
   - Both the blue dimensions Product Name and measures SUM(Profit) are discrete. Discrete fields add headers to the view and the values are considered as finite.

   Example 1:



Quantity field is considered as a measure actually. In this example, in the left visualization we had it as a continuous dimension and so it created a horizontal axis along the bottom of the view. And in the right visualization we had quantity filed as a discrete dimension. Here it created horizontal headers instead of an axis. Sales field is continuous and so it created a vertical axis in both the visualizations.There is no aggregation for the quantity field, this indicates that it is a dimension. Fields from the dimension pane are discrete initially, but the numeric or date dimensions can be changed to continuous fields. But the dimensions containing boolean or string values cannot be continuous.

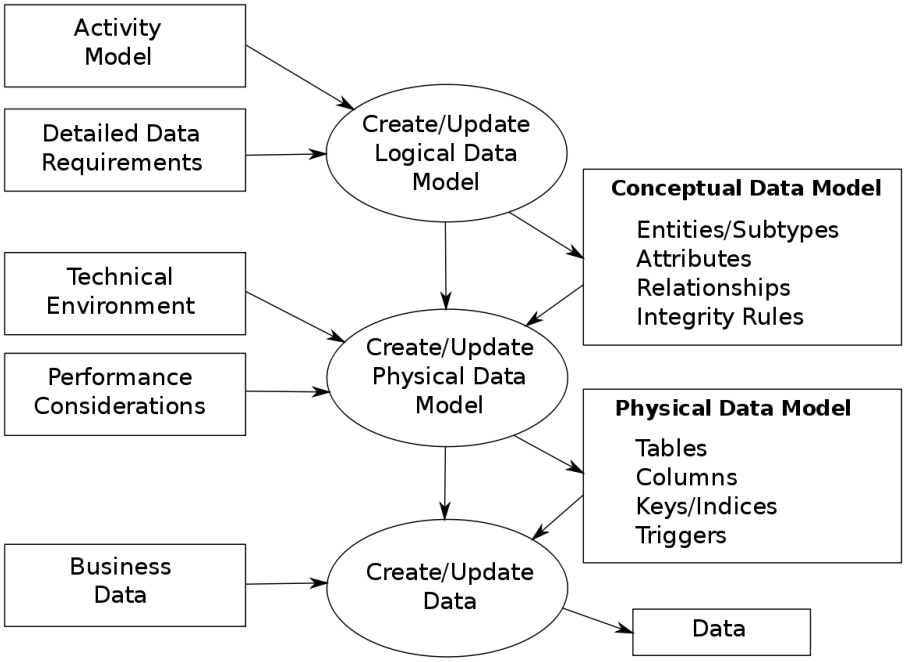
Example 2:

In this example 2, we can see that, on dragging a dimension field Category to columns, Tableau created column headers. And on dragging a measure field Sales to columns, the view will contain a continuous axis as discussed. Apart from that we can see that there is no aggregation applied for the category field because it is a dimension by default. And sum aggregation is applied to the Sales field because it is a measure.

Example 3:

In this example 3, we can see how the dimensions will affect the level of detail in the view. It refers to how granular the data is. In the left visualization we have only Category field in the Columns and we had 3 marks, but after adding another dimensional field Sub-Category to the columns in the right visualization we see that the marks have increased to 17. So, as we keep on adding dimensions to rows or columns, the number of marks in the view will increase.

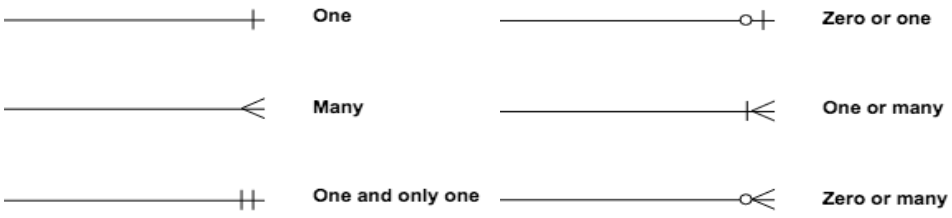
2. In the process of database designing, data modelling is the first step. Sometimes it is considered as a high-level and abstract design phase which is referred to as conceptual design. Data modelling aims to describe the data contained in the database, relationships between the data items, and the constraints on data. There are 3 different levels of abstraction at which models are developed.
   - **Conceptual Model:** This is to establish the entities, their attributes, and their relationships.
   - **Logical Data Model:** This defines the structure of the data elements and sets the relationship between them.
   - **Physical Data Model:** This describes the database-specific implementation of the data model.



The 3 basic tenants of Conceptual Data Model are:

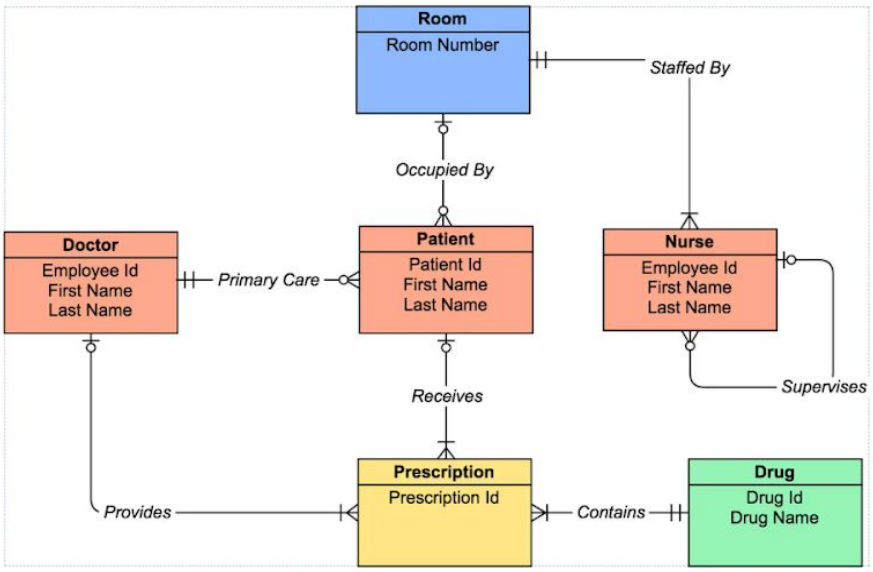
- Entity: A real-world thing
- Attribute: Characteristics or properties of an entity
- Relationship: Dependency or association between two entities

Relationship involved between the entities are as shown,



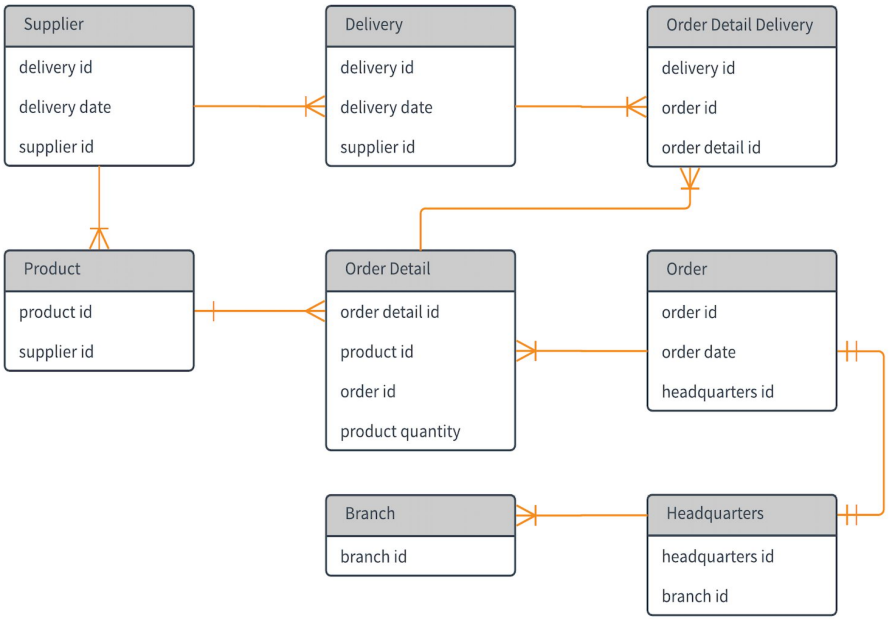
Example1:

In this example we have Room, Doctor, Patient, Nurse, Prescription, and Drug as entities. Room Number is the attribute of the Room entity, and Employee Id, First Name, Last Name are the attributes of the Doctor entity, similarly respective attributes are listed within each entity.



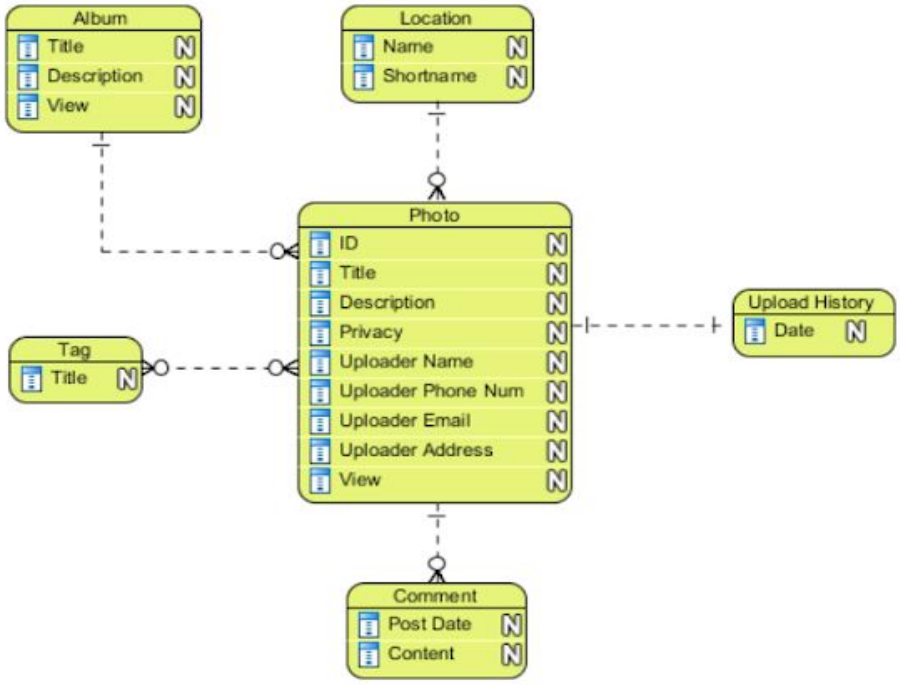
We can see that the relationship between the Doctor entity and Patient entity is one and only one on one side and zero to many on another side. Similarly other relationships are represented.

Example2:



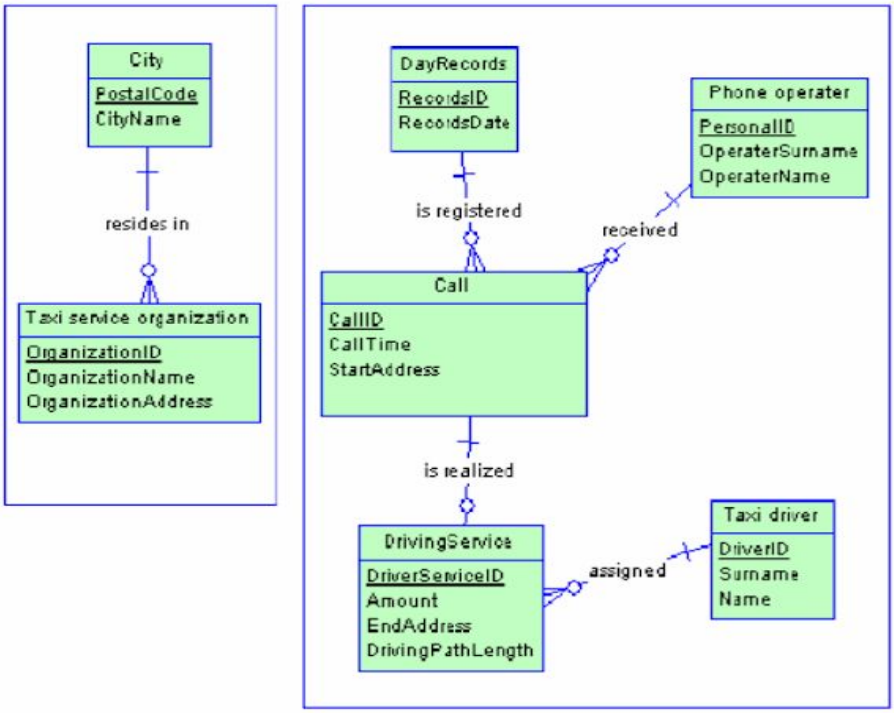
In this example, 8 entities - Supplier, Delivery, Order Detail Delivery, Product, Order detail, Order, Branch, and Headquarters. Each entity has around 1 to 4 attributes and all the entities are connected to each other based on the relationships established between them and most of them have one or many on one side of the relationship. All the entities have an ID attribute associated with them.

Example3:



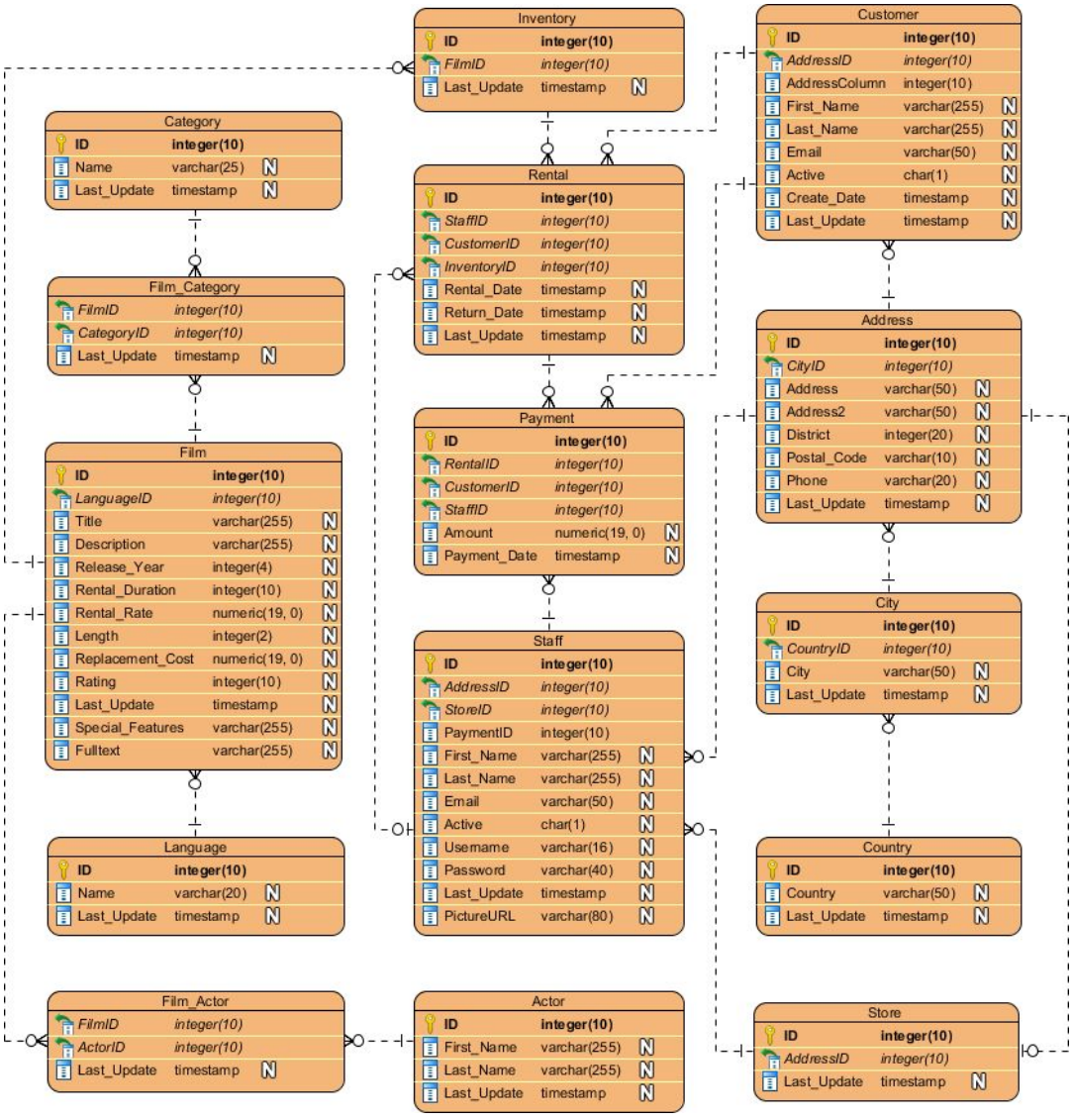
We have 5 entities in this example - Album, Location, Tag, Photo, Upload History, and Comment. Photo entity is clearly the central one which is related to all other entities with some relationship in the data model. And all entities other than Photo have no other relationship associated apart from the one with Photo entity. Their attributes are listed within each entity.

Example 4:



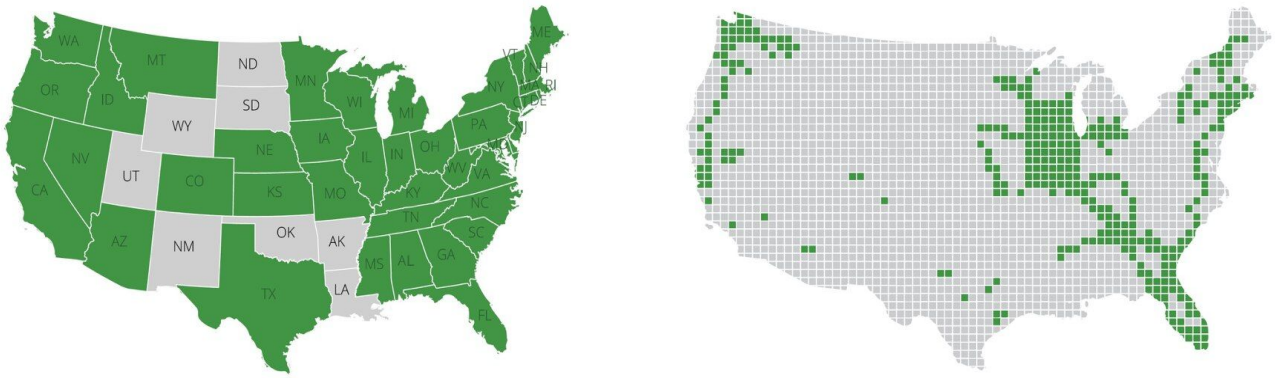
In this conceptual data model, we have two sets which are not related in any way. One contains the entities - City and Taxi service organization, and the other one consists of the entities - DayRecords, Phone operator, Call, DrivingService and Taxi driver. In both the sets we can see that all the relationships have the same property - one side of the relationship is one, and the other side is zero or many (except for one relationship with the DrivingService entity).

Example 5:



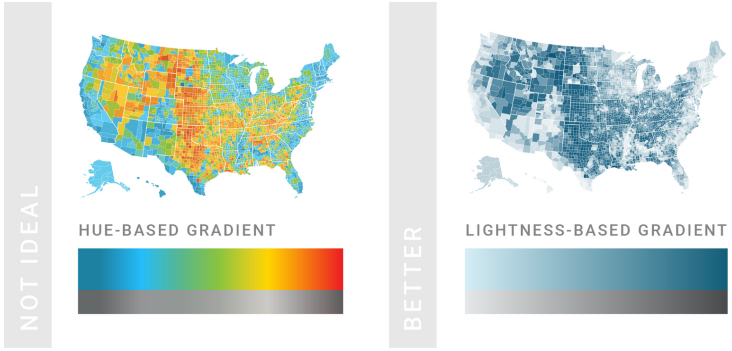
This data model is a bit complex one with 15 entities, with each entity having a considerable number of attributes. All entities are associated with 1 to 4 relationships. And most of the relationships have zero or many on one side. Except the Film_Category and Film_Actor entities, all the other 13 entities have an ID column associated with them.

3. The marketers or the analysts need to make the most out of the data, deliver the right message and should ensure that visualization will let the data tell the story. Consider the below example on how two narratives can be extracted from the exact same dataset. These two visualizations represent the places a person has traveled to in the United States. Looking at the first visualization, the person can conclude that “I have been all over the country”. But in the right visualization, obtained by removing the state lines and dotting the exact locations, the person can conclude as “I guess I haven’t been all over the country”.

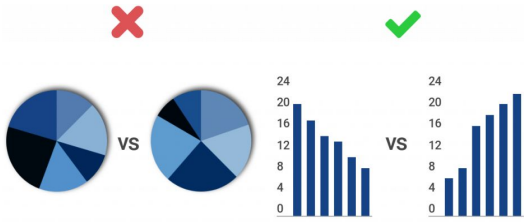


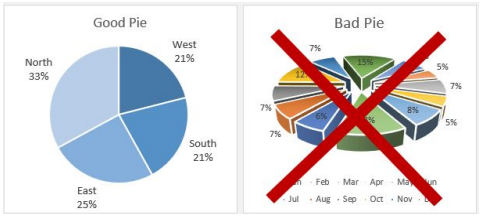
This example clearly illustrates that “While the data itself isn’t malleable, the message is”. Avoiding the below pitfalls that are most common in data visualization can help us in conveying the right message clearly.

1) **Color Abuse:** We shouldn’t overdo the color in data visualizations. It has its place, but then wrong color can lead to confusion, or even more worse, misinterpretation. Instead of relying on color alone to convey the meaning, consider the color blind and use the shapes and colors that are easiest for viewers to grasp the information.

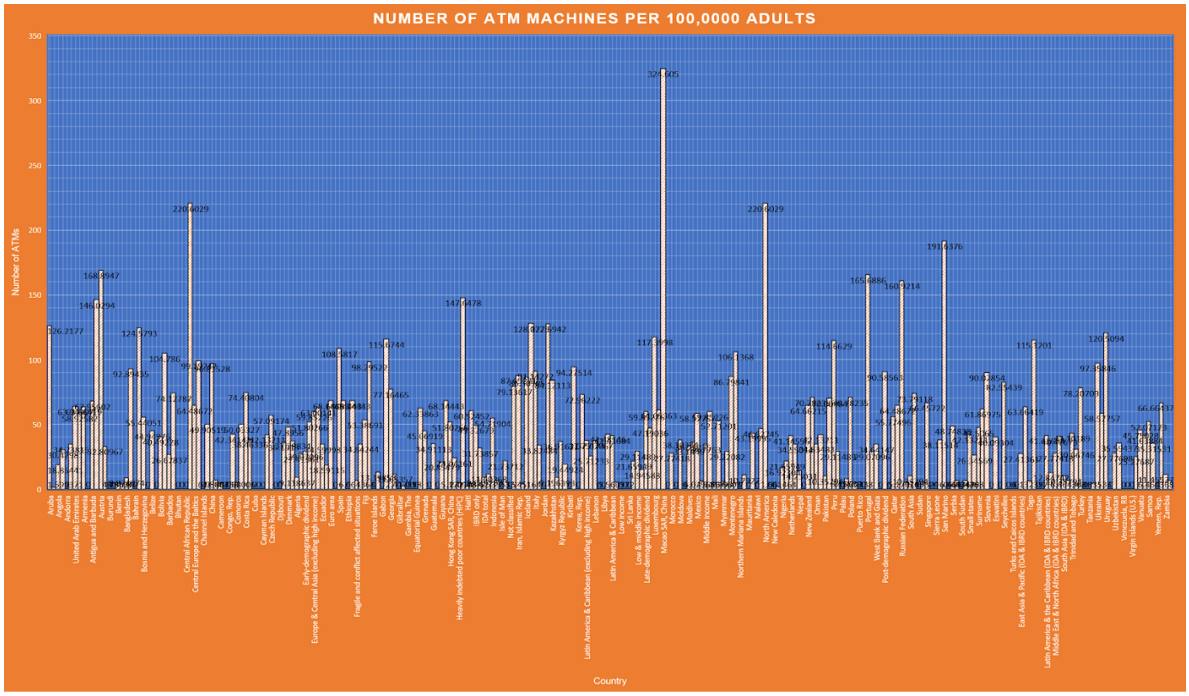


2) **Misuse of Pie Charts:** On trying to squeeze too much information into pie chart, the main and big picture will get lost. Too much detailing in the pie chart will leave the viewers feeling confused and unsatisfied. It's an awkward way for comparison, and so, avoid using pie charts side by side. It is best to use for limited dimensional values where each slice of the pie can be easily distinguished. Use them to compare parts of a whole but not to compare different sets of data. For easy comparison, order the slices from smallest to largest.





3) **Visual Clutter:** Too much information will defeat the purpose of clarity. And making discoveries in a cluttered visualization is similar to finding a needle in a haystack. Adding unnecessary elements to a visualization will obscure meaning and lead to inaccurate conclusions. If the visual is looking clustered, try a different format. Too many objects are distracting, so keep the visualization or dashboard simple by limiting the objects to eight or less.



Number Of ATM Machines Per 100,000 Adults By Country

Country Name	2018
Macao SAR, China	324.61
North America	220.60
Canada	220.60
San Marino	191.64
Austria	168.89
Portugal	165.69
Russian Federation	160.92
Croatia	147.65
Australia	146.03
Israel	128.07
Japan	127.59
Aruba	126.22
Bahamas, The	124.58
Uruguay	120.51
Luxembourg	117.40
United Kingdom	115.67
Thailand	115.12
Peru	114.66
Spain	108.58
Mongolia	106.14
Brazil	104.79
Switzerland	99.19
France	98.30
Ukraine	97.36
China	96.82
St. Kitts and Nevis	94.28
Bulgaria	92.89
Italy	91.14
Post-demographic dividend	90.59
Slovenia	90.03
Iran, Islamic Rep.	87.69
Montenegro	86.80

4) **Poor Design:** Just because a visualization is elegant to look at, doesn't mean that it is effective. Effective visualizations should incorporate best design practices to strengthen the communication of data. Before building the visualization, it is worth enough to think on how users will navigate through our charts and answer the following questions - how do they need to see it, what do they need to see, what additional context they may require and how can they access it?

In this example, Graphics add no value. And to get the information from this title, we need to read the text associated with each graphic.

5) **Bad Data:** Good visualizations will start with good data. If the visualization is revealing unexpected results, maybe we are the victim of bad or improper data. We shouldn't let our visualization become the scapegoat for inadequate or substandard data. Before presenting the data, identify the issues with data using charts and address them. Below is the example for a poor visualization design.

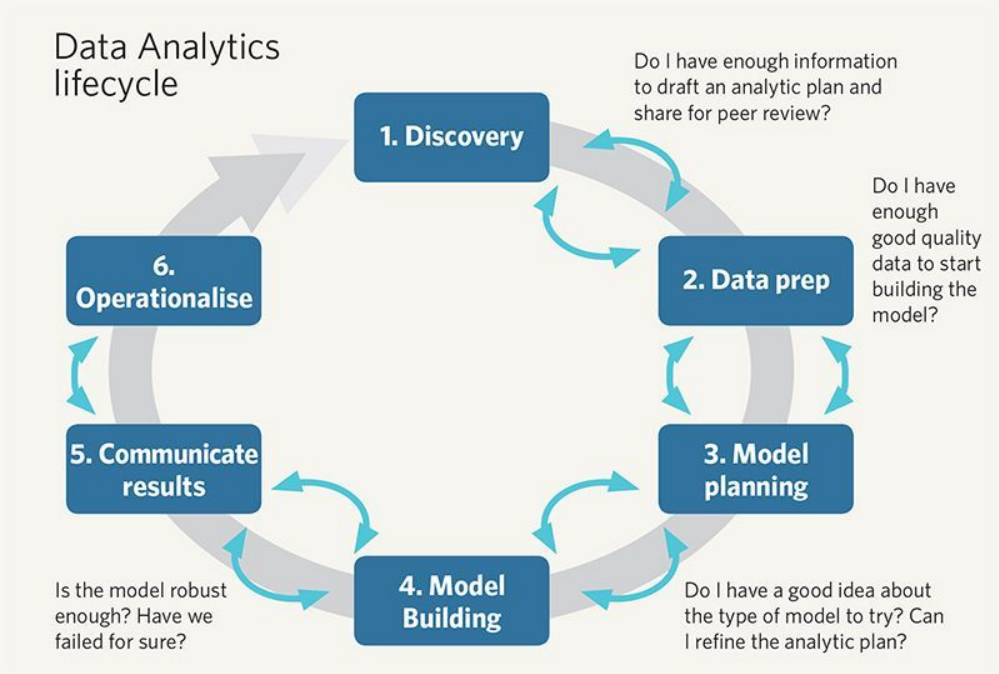
Sticking onto these five rules, we will be on our way to create effective visualizations and deliver the right message to the audience.

PART II: Value of Data Visualization

Data Analytics encompasses the below six phases. These phases of the life cycle are iterative with forward and backward and even overlapping movement sometimes when new information is uncovered. Stakeholders of a data analytics project are:

- Business User/Customer: Understands the domain area
- Project Sponsor: Provides requirements, resources and support
- Project Manager: Ensures meeting objectives and the progress of the project
- Business Intelligence Analyst: Provides business domain expertise based on the deep understanding of data
- System Administrator: Works on setting up and maintaining the systems (creates DB environment)
- Data Engineer: Provides technical skills and support, supports analytic sandbox, assists data extraction and its management for harvesting, storing, retrieving, and processing data.
- Data Scientist: Provides analytic techniques and modeling

This life cycle lays out the framework for best practices from the genesis of the project till the completion of the project.



1) **Discovery:** In this phase the team must investigate and analyze the problem, develop understanding and context, learn about needed and available data sources, and should formulate initial hypotheses that can later be tested with data. The five main activities to be performed during this phase: Identify data sources, Capture aggregate data sources, Review the raw data, Evaluate data structures and tools needed, Scope the sort of data infrastructure

needed for this type of problem. With the best interactive visualization tool, we can ask questions through direct interaction, and can use multidimensional visualizations to see the relationships between different variables for data discovery.

2) **Data Preparation:** This phase includes steps to explore, preprocess, and condition data prior to modeling and analysis. Pay attention to the field/variable with many missing values. The raw data is cleaned and transformed to the required format for easier understanding, better analysis, and better decision making which altogether is a process called Data Wrangling. At this stage as we have data in a variety of sizes and shapes, wrangling of data is mostly done with semi-structured data. This should be done by the experts who know the dataset well. By visualizing the data, we can have an overview of data and preliminary insights into it.

3) **Model Planning:** In this phase we determine the techniques, methods, and workflow it intends to follow. We will explore the data using visualization and subsequently select key variables and the most suitable models based on the structure of data (Structured data, Unstructured data, or Semi-Structured data). Once we get acquainted with the data, we refer to the hypothesis developed in the discovery phase. Common tools for model planning are: R, SQL Analysis, SAS/ACCESS.

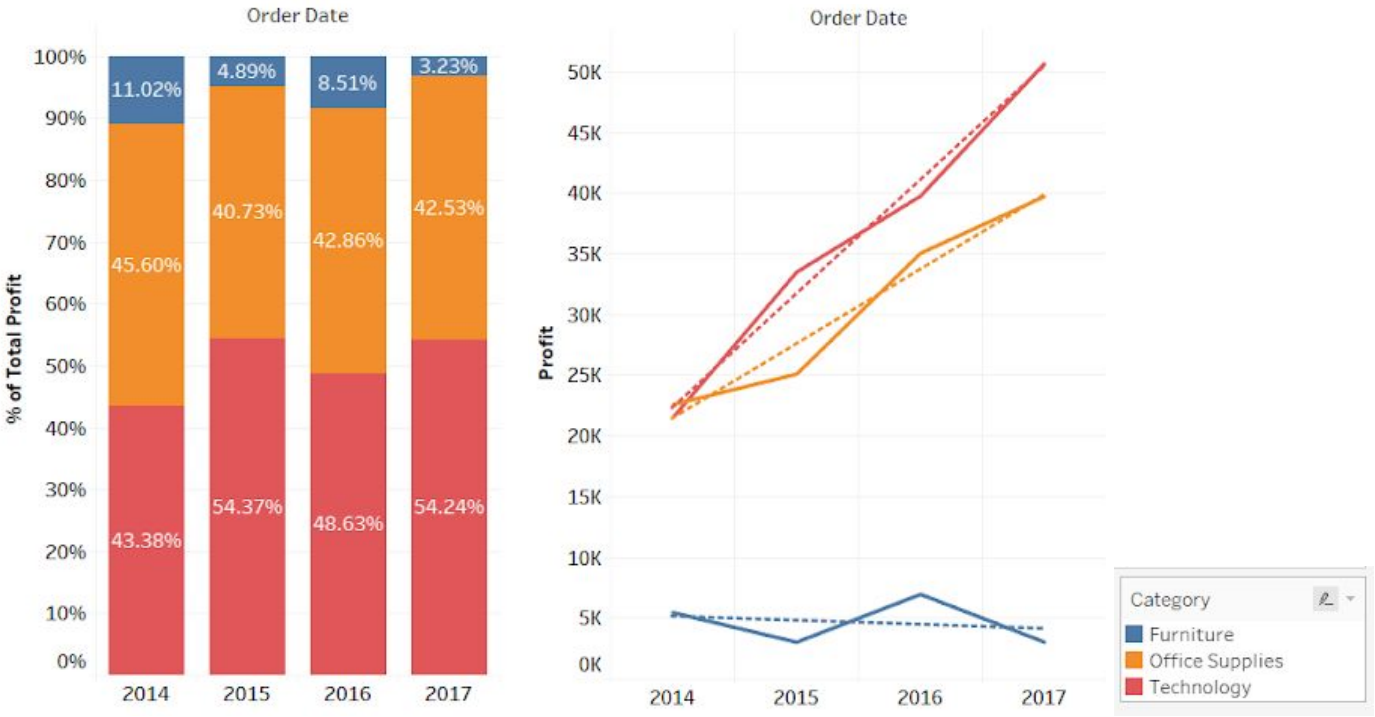
4) **Model Building:** In this phase we need to develop datasets for training, testing, and production processes. The analytical model developed is trained on training data and tested on testing data. We should also check if the existing tools are sufficient for running the models or need a robust environment for executing models and workflows.

5) **Communicate Results:** During this phase we will compare the outcomes of the model to the established criteria for success and failure. We consider how best to articulate the outcomes and findings to stakeholders and various team members, taking into account assumptions, warnings, and any other limitations of the results. We need to identify key findings, quantify business value, and develop a narrative to summarize and convey the findings to stakeholders.

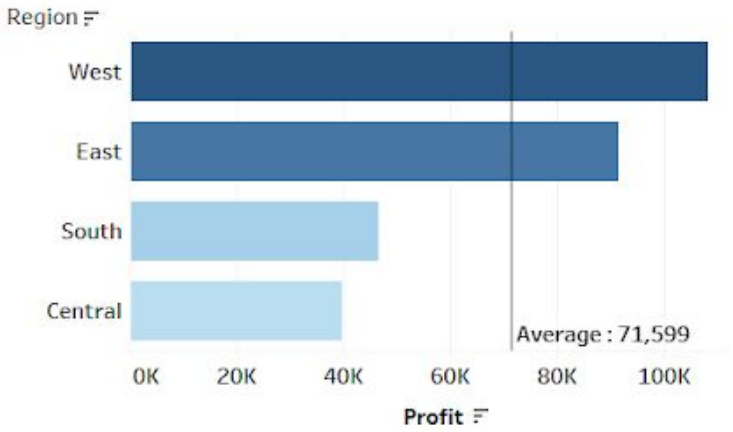
6) **Operationalise:** In this phase, we communicate the benefits of the project more broadly and set up a pilot project to deploy the work in a controlled way before broadening the work to a full enterprise or ecosystem of users. During this approach we will learn about the performance and constraints of the developed model in a production environment on a small scale and make necessary adjustments before a full deployment. And the deliverables include final reports, briefings, code, and technical documents.

PART III: Data Visualization, EDA & Tableau

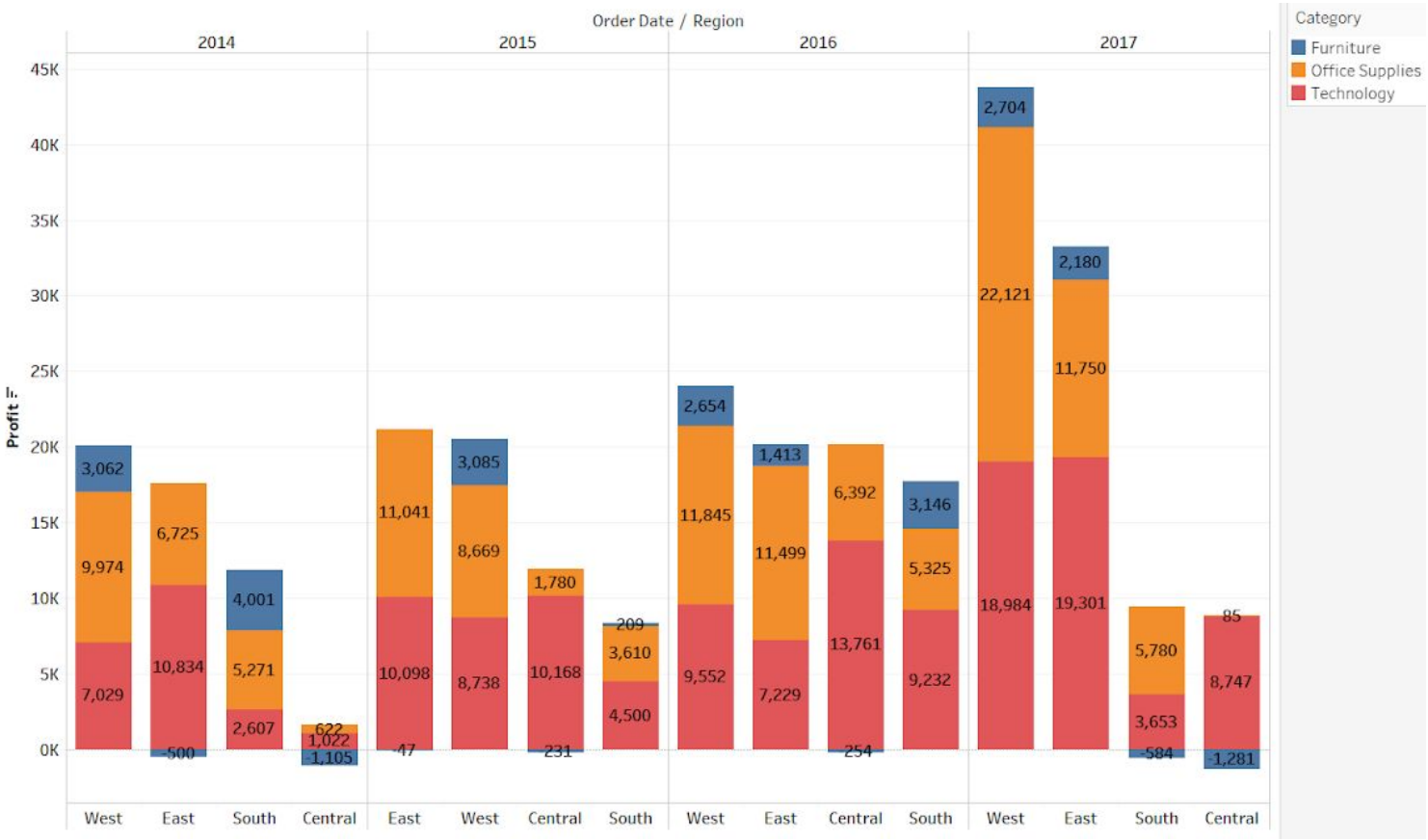
In the below left visualization we can see the percentage of total profit of each category - Furniture, Office Supplies, and Technology in a given year. Office Supplies has a relatively steady percent of total profit. In all the years, it is clearly evident that most of the profit comes from the Technology and Office Supplies categories. Comparatively, Technology category has a bit higher profit percentage than Office Supplies during all the years. Furniture has a very low percentage of profit every year, and it's almost decreasing during the period. In the below right visualization, profit is trending upward for both the categories Office Supplies and Technology, and downward for Furniture during the period.



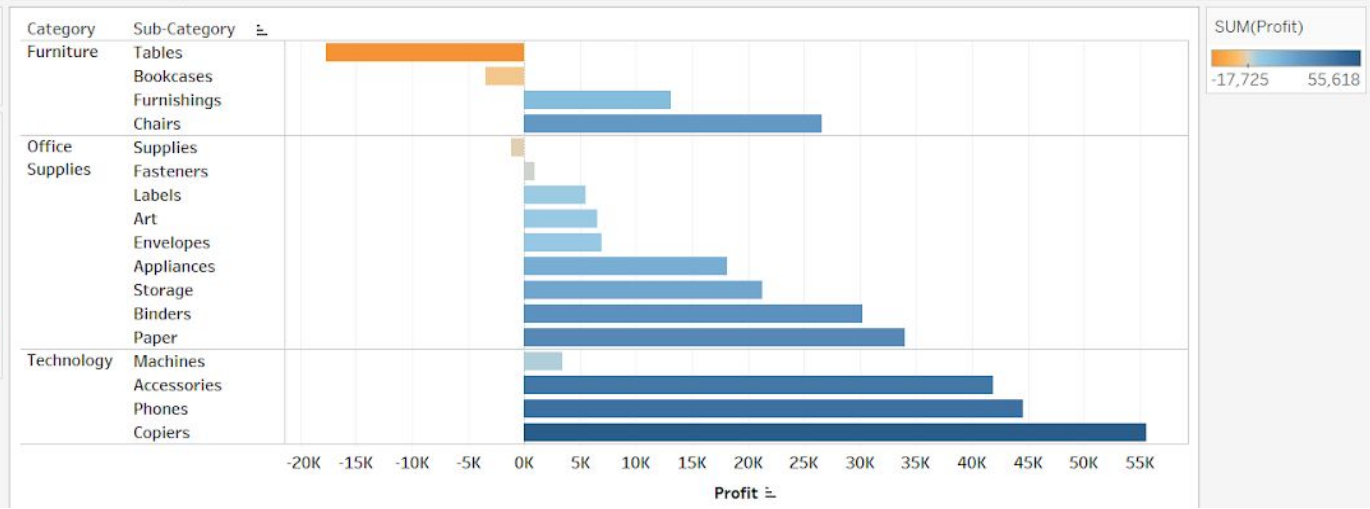
From the below visualization on Profit over Region with a reference line on Average Profit, it's clear that West and East regions have noticeably more profit compared to South and Central regions.



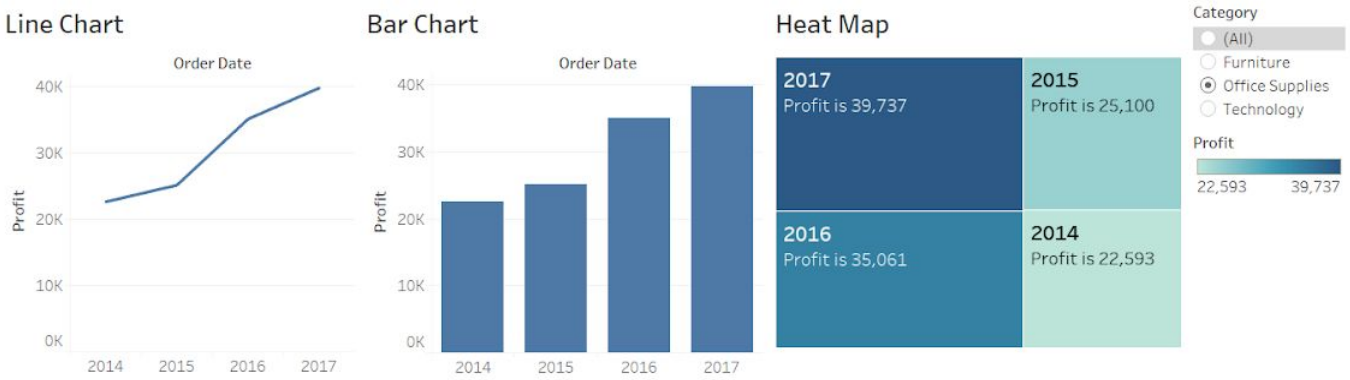
Below visualization shows trends of Profit over Order Date and Region with level of detail increased with Categories - Furniture, Office Supplies, and Technology. During all the years, West and East regions have made more profit than the South and Central regions. Only the Furniture category had negative profits. In the Central region, the Furniture category had negative profits during the whole period. The East region had profit of (-500) during 2014 and (-47) during 2015 which decreased gradually and made positive profit of 1413 during 2016 and 2180 during 2017. And the South region had profit of (-584) for the first time in the year 2017.



The below visualization is showing the profit for all categories drilled-down to their sub-categories. Tables have noticeably less Profit of (-17,725) among all the subcategories. Copiers subcategory in the technology category have the highest profit of 55,618 among all others. All the subcategories in the Technology have positive and relatively high profit. Except for the Supplies subcategory from the Office Supplies category, all others have positive profit. Tables and Bookcases have negative profits in the Furniture category. Profits distribution is shown with a diverging orange-blue color. Orange colored are the ones having negative profits, and blue colored are the ones with positive profits.



Looking at the Dashboard - Chart Types 1 provided below, with a bar for each year, we can show the profit for each year as a bar either extending towards the positive side of the vertical axis or down towards the negative side of the vertical axis in a bar chart. It is the same with line charts, but then here we have a continuous plot over time. Coming to heat map, it uses a dark-to-light color scale to display or focus on the year that has more profits with a dark color and the year that has less profits with a light color. But it is not a good visualization to consider when we have many years because it is hard to get the insight for comparing profits over time. So clearly, it's either a bar chart or line chart. But particularly if any changes are small, line graphs are comparatively better to interpret than bar graphs. Line graphs even makes the overall trends very clear. And it is better for representing trends over time or any other measure having a logical progression of values. But in this scenario when we dig deep to analyze and filter out the questions we had, there are so many attributes in the view and the line graph will be so cluttered and confusing. So, bar charts are relatively best to visualize profits over time with all these attributes.

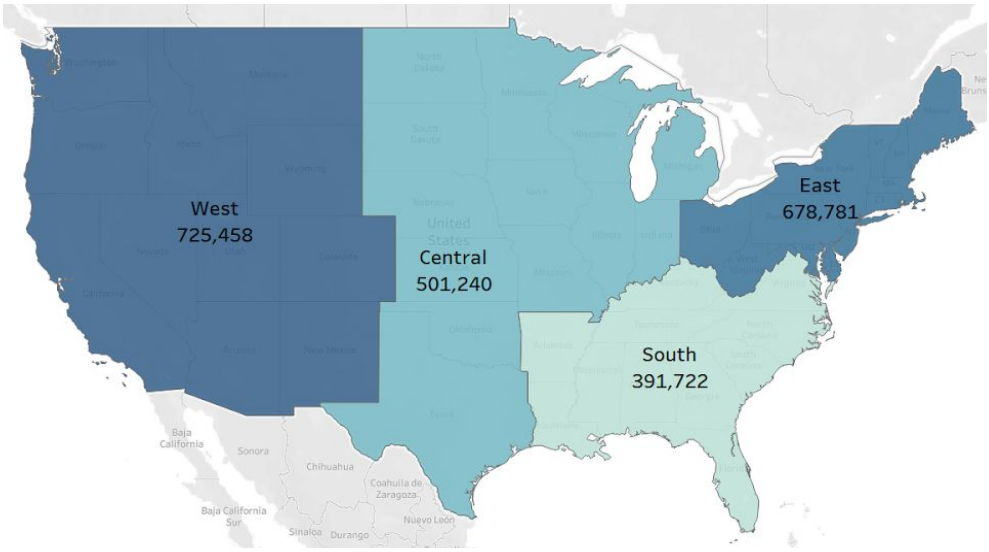


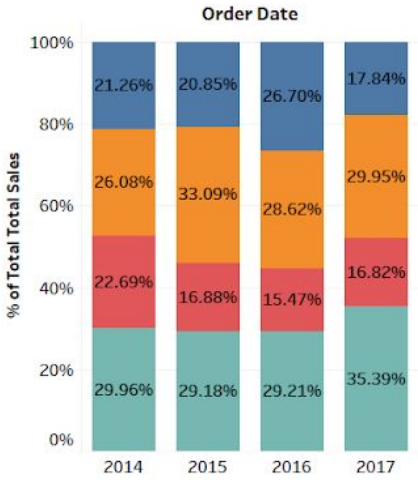
Dashboard - Chart Types 1

We could even drill down the order date from year to quarters and even further to months and represent for deeper analysis. Progression from left to right will show us how the profit has changed in successive quarters.

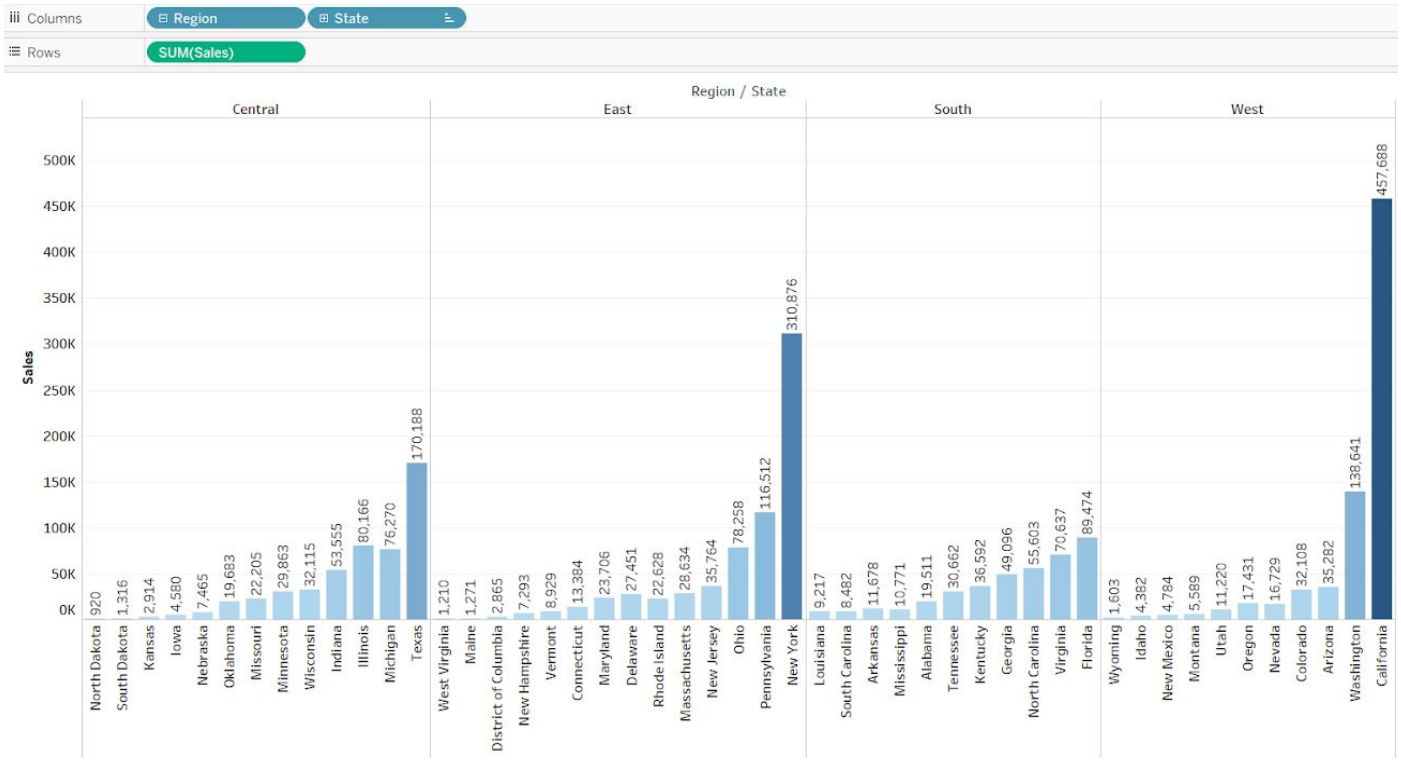
PART IV: Data Visualization, EDA & Tableau

The two visualizations below show the total sales in all the four regions - West, Central, East and South. It's clear that the company's business is doing the best with the highest sales figure of 725458 in the West region, then 678781 in the East region, then 501240 in the Central region, and last 391722 in the South region. In the second graph we can see the percentage of sales of each region during the period. In the next visualizations we can see how the sales vary for all the states, and categories and their subcategories as well.

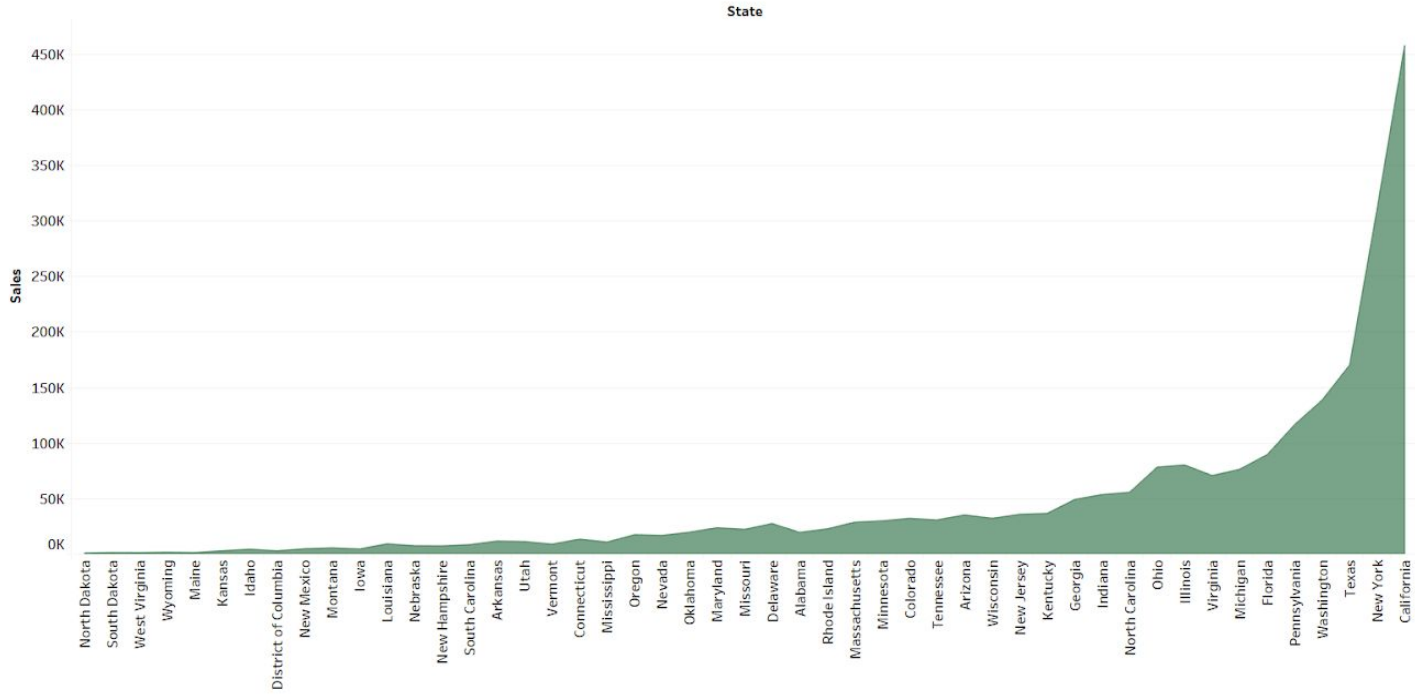




With a location hierarchy in the order - Region, State, and City, on drilling down from the Region to their States, observed the States that are doing best with high sales in each of the four regions in the below visualization. In the Central region, Texas has the best sales of 170188 and North Dakota has the lowest sales of 920. In the East region, New York has the best sales of 310876 and West Virginia has the lowest sales of 1210. In the South region, Florida has the best sales of 89474 and Louisiana has the lowest sales of 9217. In the West region, California has the best sales of 457688 and Wyoming has the lowest sales of 1603 as shown below. We can add a filter for Region and State fields, so that it's easy to compare the Sales when we drill down a level further to City.



Below is the area visualization of total sales against each State without a division of Region to capture the order of the states that have the strongest sales figures for the company. With the YEAR(Order Date) as filter, we can change over the period from 2014 to 2017. California and New York have always stayed as the top two states with strongest sales of 457688 and 310876. Overall, in all the years, North Dakota and South Dakota are the two states with lowest sales of 920 and 1316.



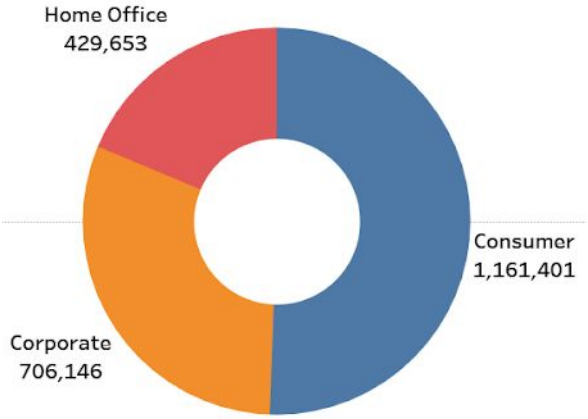
Below visualization shows the total sales for all the sub-categories within all States for each Region. From this visual we get to know that the state Wyoming in the West region has sales only in the Furniture category and that too, only in the Chairs sub-category. That could be the reason for its lowest sales among all states in the West region or may be there is some missing data. Same is the case with the West Virginia state in the East region with sales only in three subcategories - Tables, Binders, and Paper.

		Category / Sub-Category																
		Furniture				Office Supplies									Technology			
Region	State	Bookcases	Chairs	Furnishin..	Tables	Applianc..	Art	Binders	Envelopes	Fasteners	Labels	Paper	Storage	Supplies	Accessori..	Copiers	Machines	Phones
Central	Illinois	4,283	14,563	2,878	6,551	975	930	4,539	384	141	225	3,456	9,080	178	5,536	5,920	3,756	16,772
	Indiana		6,463	1,514	3,519	4,160	389	4,012	890	8	276	1,880	4,120		2,279	18,500	84	5,460
	Iowa		1,408	49	1,185		147	248	13	46		316	13					1,154
	Kansas			111		82	163	612		24	19	303	394	358	92			757
	Michigan	810	13,878	1,916	5,717	4,324	1,005	22,822	310	110	881	2,011	6,187	74	4,933	1,150	3,411	6,731
	Minnesota		6,079	232	1,300	2,844	103	12,470	31	43	161	320	3,398	37	1,520	550		775
	Missouri	213	61	941	1,722	3,670	240	1,876	71		15	302	1,792	4,217	1,022	5,500		565
	Nebraska		564	1,381		501	19	128		58	14	333	1,165	17	240			3,046
	North Dakota						182	26		7			705					
	Oklahoma	342	3,963	1,445	2,534	1,491	59	445	407		64	197	2,345	22	1,817			4,551
	South Dakota	142	183			416		115		7	35			26	393			
	Texas	14,493	26,572	3,767	15,761	2,408	2,370	9,043	2,531	332	584	6,983	15,724	4,517	11,329	5,640	19,546	28,590
	Wisconsin	3,874	11,495	1,020	868	2,711	160	588		2	178	1,392	1,007	22	4,795			4,003
East	Connecticut	1,683	2,782	457	252	1,479	192	372	11		33	788	2,428	115	658			2,133
	Delaware	2,052	1,560	637	510	140	535	2,434	299	26	74	441	4,179		562	10,500	66	3,435
	District of Columbia		1,268	79			57	10				72					1,380	
	Maine			109		102		38				260			323			438
	Maryland	882	4,330	3,148	790	1,589	434	3,225	337	27	172	1,039	3,542	25	1,124		465	2,577
	Massachusetts	1,874	4,246	834	3,966	3,348	226	2,105		51	166	1,774	4,289	30	897	1,200		3,630
	New Hampshire			833	1,053	33	266	763		15	41	285	337	28	2,812			825
	New Jersey	2,276	2,662	951	418	4,204	743	3,530	154	54	66	1,422	4,669	114	1,024		9,900	3,578
	New York	22,777	46,634	10,182	13,779	13,276	2,984	32,457	1,777	287	1,291	8,457	26,928	2,564	19,598	17,200	43,184	47,503
	Ohio	2,078	10,146	4,089	7,887	4,808	840	1,917	562	205	162	2,146	7,264	479	8,223	3,840	8,978	14,635
	Pennsylvania	5,231	18,724	7,348	8,052	4,663	1,152	6,266	1,234	155	598	2,378	11,785	6,710	7,299	13,080	2,134	19,702
	Rhode Island	563	3,194	403	1,758	2	49	292				570	4,626	695	2,210	7,400		865
	Vermont	4,405	715			543	8		2			92	1,564		305			1,295
	West Virginia				673			89				448						
South	Alabama		3,916	40	2,376	208	301	467	106	4	506	883	1,249	485	2,323	900	3,040	2,706
	Arkansas	639	1,673	573	302		78	2,890	312	43	7	806	430		1,788			2,137
	Florida	2,587	11,169	3,746	5,485	3,558	711	3,690	211	49	386	2,426	7,419	1,068	5,862	1,440	24,574	15,092
	Georgia	3,794	1,858	1,805	865	3,432	527	11,672	447	15	245	2,957	7,381	40	3,658	3,000	396	7,005
	Kentucky	262	9,234	1,788	843	1,336	351	3,696	105	76	363	1,059	4,822	86	2,976	1,400	3,080	5,114
	Louisiana	388	802	182	1,591	17	130	629	62		11	104	805	1,666	1,668		480	683
	Mississippi	505	866	430	2,516	616	607	967	91	35	15	725	576		2,010			813
	North Carolina	959	2,724	1,791	9,682	4,138	603	2,562	431	66	229	2,594	3,438	302	3,507	960	12,621	8,996
	South Carolina		2,538	540			35	515	319	43		332	2,567		139			1,453
	Tennessee	518	5,075	1,699	6,214	1,915	436	5,148	177	87	115	1,060	3,311	98	411		740	3,656
	Virginia	1,248	5,322	4,709	14,042	4,306	876	4,794	1,084	85	475	1,206	3,769	4,575	2,936	1,600	8,960	10,650
West	Arizona	519	6,499	2,512	3,995	774	1,389	2,185	215	101	27	2,441	2,590	283	3,396		966	7,389
	California	27,559	63,079	19,996	45,430	24,176	5,496	28,473	3,237	480	2,899	16,758	45,112	15,721	37,255	24,560	29,492	67,965
	Colorado	1,584	8,245	1,958	1,457	1,367	211	509	76	45	18	883	3,387	1,403	2,288	440	3,314	4,925
	Idaho		338	432	1,825	228		72				173	477		90			748
	Montana			64				502			15	41	1,304		217	3,000		446
	Nevada	142	674	463	3,355	96	214	4,897		5	15	318	1,326	86	230		2,396	2,511
	New Mexico		1,114	587			277	126	28		18	626	280	30	1,164			534
	Oregon	357	3,841	468	1,672	536	462	281	50	20	902	953	1,991	77	1,160		210	4,451
	Utah	2,700		165	1,957	1,090	100	760		10	30	1,152	902	44	195	1,500		615
	Washington	3,144	16,387	3,427	25,062	1,969	1,064	18,156	513	263	1,154	3,319	13,163	483	15,119	20,250	6,066	9,102
	Wyoming		1,603															

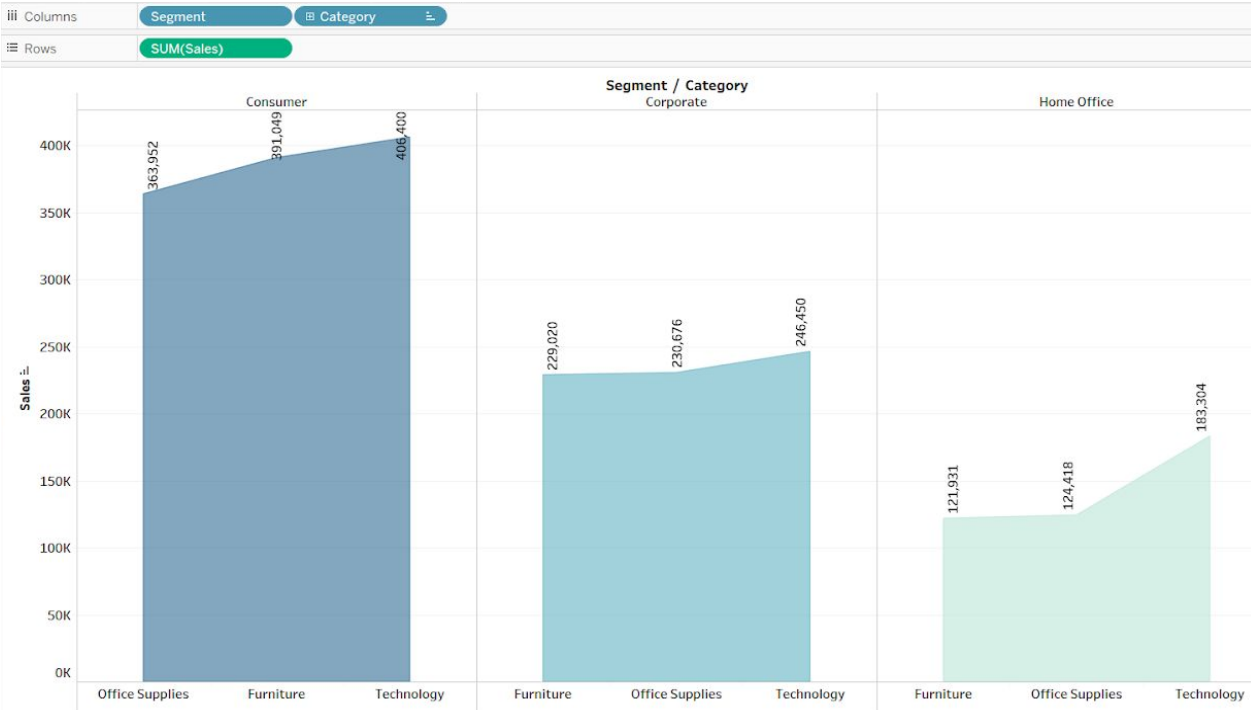
In this scenario, as we are supposed to look at how the business is doing in the four regions and in the states within, map visualization would be the best one to visualize among the regions because we have only four. But to drill down further to states within each region, bar graph will be suitable for easiest comparison and to avoid confusion. To see for any missing data or to find the states that have sales in very few sub-categories, the last visualization was best. And so, on whole, bar graph is the best one to compare without confusion.

PART V: Data Visualization, EDA & Tableau

Below is a donut chart showing the total Sales for the three segments of customers - Consumer, Corporate, and Home Office. It's evident that the Consumer segment accounts for the majority of Sales of 1,161,401 which is more than 50 percent of the sales (50.56%). The Corporate segment accounts for 30.74% of sales and the Home Office segment accounts for 18.7% of sales. And so, the company should focus most on the Consumer segment of customer.

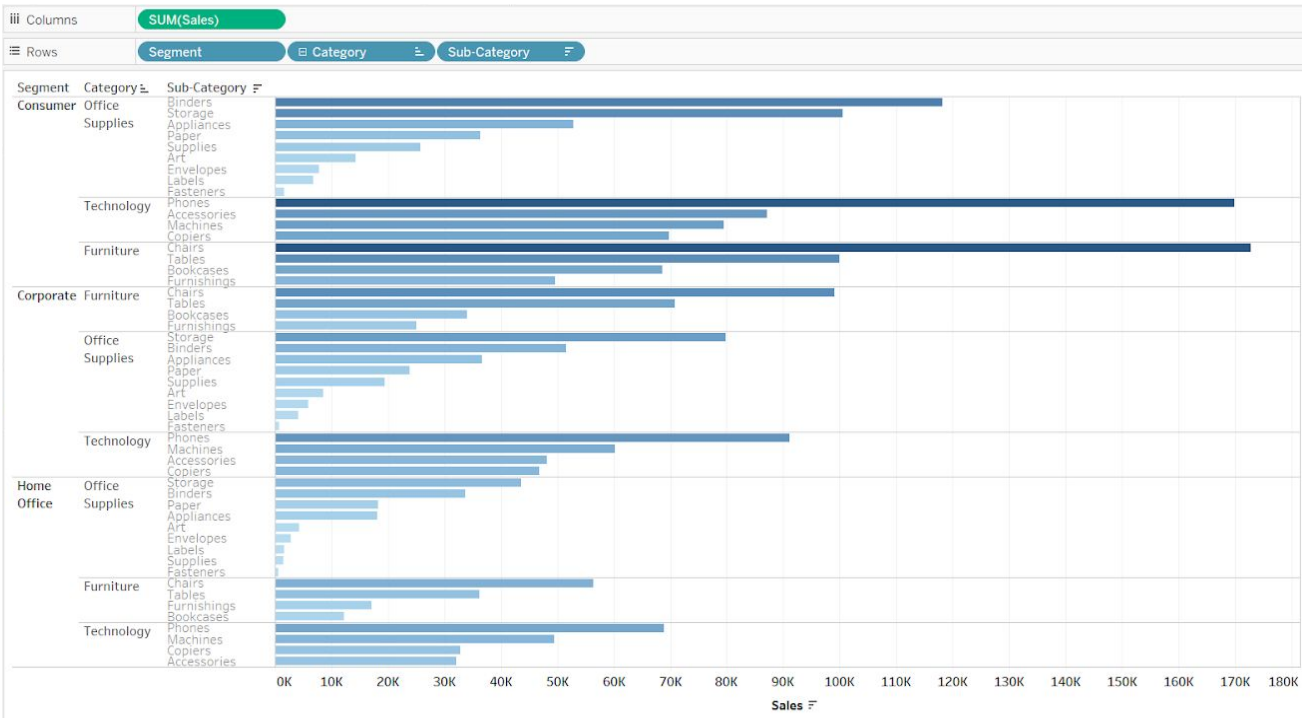


The area visualization presented below shows how the Sales vary among the Categories for all the segments of customer. Category can be drilled down to Sub-category for further details. The technology category has the highest sales in all the three segments, 406,400 in the Consumer segment and 246,450 in the Corporate segment 183,304 in the Home Office Segment.

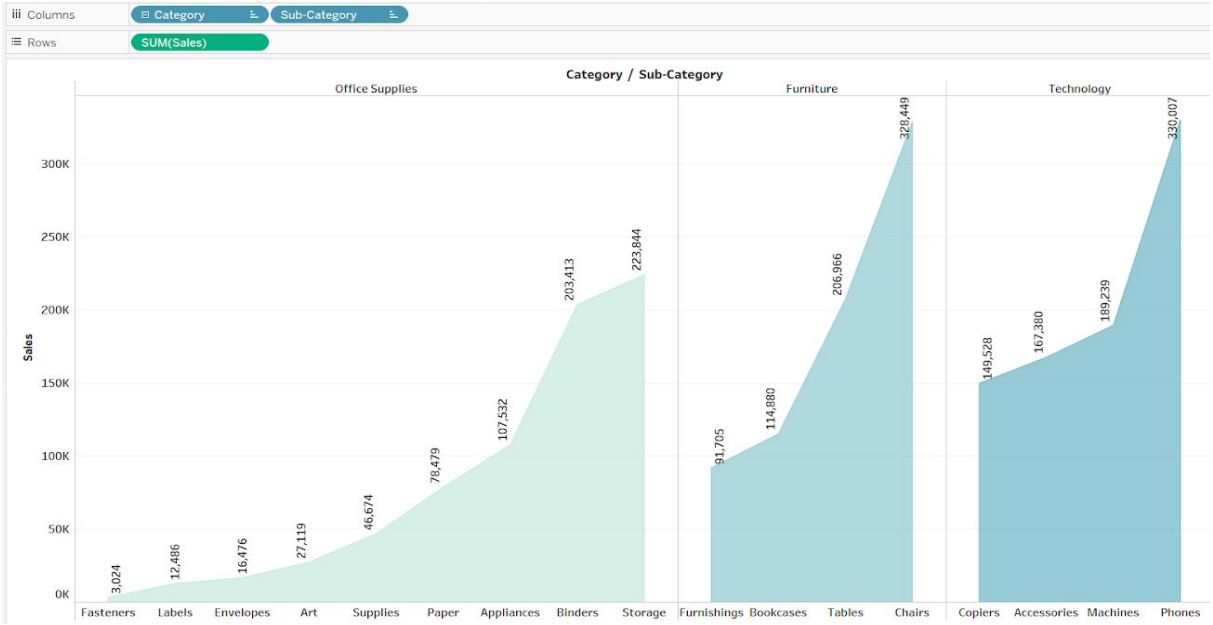


The below bar chart shows the sales among all the segments drilled down to categories, and all the categories drilled down further to their sub-categories. Chairs sub-category from the Furniture

category and Phones sub-category from the Technology category in the Consumer segment accounts for the highest sales of 172863 and 169933 among all the sub-categories. Fasteners in all the three segments, Supplies and Labels in the Home Office segment are the subcategories with lowest sales.



The area visualization shown below plots the total sales for all the sub-categories in each category. The marks on top of the area graphs represent the exact total sales value for each of the sub-category. Phones sub-category has the highest sales of 330,007 in the technology category. Chairs sub-category has the highest sales of 328,449 in the furniture category. Storage sub-category has the highest sales of 223,844 in the office supplies category.



For this scenario, the donut chart was used to show how the three segments account for the sales. As we have only three segments of customers, this was the suitable one to convey a rough idea, in the case of more segments donut or pie chart would make it very difficult to get the insight quickly. Drilling down further to categories and sub-categories, we used the area chart which is a kind of combination of line chart and bar chart visualizations. In this way we can see the changes easily and can easily capture their contribution to total sales in a glance. And so, Area chart visualization suits best for this scenario.

References :

1. https://help.tableau.com/current/pro/desktop/en-us/datafields_typesandroles.htm
2. https://evolytics.com/blog/tableau-fundamentals-dimension-vs-measure/#:~:text=According%20to%20Tableau's%20Knowledge%20Base%2C%20a%20dimension%20is%20a%20field,categorical%20information%20as%20a%20dimension.&text=Generally%2C%20the%20measure%20is%20the,and%20dice%E2%80%9D%20the%20number%20by.
3. https://online.visual-paradigm.com/knowledge/visual-modeling/conceptual-vs-logical-vs-physical-data-model/
4. https://www.visual-paradigm.com/guide/data-modeling/what-is-data-modeling/
5. https://mschermann.github.io/data_viz_reader/how-to-run-a-data-visualization-project.html
6. https://theindex.generalassemb.ly/five-data-visualization-mistakes-and-how-to-avoid-them-74e3c595f5f9
7. https://medium.com/@gp_pulipaka/an-approach-to-machine-learning-anddata-analytics-lifecycle-e79c0ad55005