PROPOSAL PROYEK 12S4056 – VISUALISASI DATA

Brazilian E-Commerce Public Dataset Visualization

Disusun Oleh:

12S17016 Heppy Maria Simanungkalit

12S17025 Evola R.A Tampubolon

12S17036 Tri Dessy Natalia



PROGRAM STUDI SARJANA SISTEM INFORMASI FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO (FITE) INSTITUT TEKNOLOGI DEL TAHUN 2020/2021

DAFTAR ISI

DAFTA	R ISI	ii
DAFTA	R GAMBAR	iii
DAFTA	R TABEL	iv
BAB 1 I	PENDAHULUAN	1
1.1	Latar Belakang	1
1.2	Tujuan	2
1.3	Pertanyaan Visualisasi	2
1.4	Manfaat	2
1.5	Ruang Lingkup	3
BAB 2 I	ISI	4
2.1	Visualisasi Data	4
2.2	Aktivitas Pengembangan Dashboard Visualisasi Data	
2.2		
2.2		
2.2		
2.2	.5 Validation	11
2.2	.6 Membangun Panel Idiom dan <i>Dashboard</i> Visualisasi dengan <i>Visu</i>12	ualization Tool
BAB 3 I	RENCANA KERJA	13
3.1	Jadwal Kegiatan	13
3.2	Team Member and Roles	14
Daftar P	Pustaka	15

DAFTAR GAMBAR

Gambar 1 Tahapan Pengembangan Visualisasi Data	5
Gambar 2 Kerangka kerja fFramework) menurut Tamara Muzner	6
Gambar 3 Tipe Dataset menurut Tamara Munzner	7
Gambar 4 Tipe Data menurut Tamara Munzner	7
Gambar 5 Data dan Dataset Types menurut Tamara Munzner	8
Gambar 6 Attribute Types menurut Tamara Munzner	8
Gambar 7 Actions dan Target menurut Munzner	9
Gambar 8 How to design vis idiom: encode, manipulate, facet, and reduce by Tamara	
Munzner	10
Gambar 9 Expressiveness Types and Effectiveness Ranks by Tamara Munzner	11

DAFTAR TABEL

Tabel 1 Jadwal Kegiatan	.13
Tabel 2 Team Members and Tasks	.14

BAB 1 PENDAHULUAN

Bab ini menjelaskan latar belakang, tujuan, pertanyaan visualisasi, manfaat dan ruang lingkup pengerjaan proyek.

1.1 Latar Belakang

Perkembangan teknologi informasi yang semakin maju memiliki dampak bagi perilaku informasi masyarakat di berbagai bidang seperti bidang pendidikan, ekonomi, bisnis, kesehatan, hiburan, komunikasi dan lainnya [1]. Terdapat juga sebuah istilah baru yang disebut dengan *Big Data* yang berisi sejumlah data besar yang perlu dianalisis untuk dapat dijadikan sebagai suatu acuan dalam pengambilan keputusan ataupun dalam memperoleh sebuah pengetahuan baru. Informasi dapat lebih mudah dan cepat untuk dikirim dan diterima melalui media *online*, dengan menggunakan internet untuk mengaksesnya. Di bidang ekonomi, perkembangan teknologi informasi dimanfaatkan untuk mendukung bagian usaha dan aktivitas transaksi bisnis seperti dalam melakukan transaksi jual-beli produk atau jasa yang dilakukan melalui media online. Internet dan alat elektronik seperti komputer, laptop, handphone dan lainnya dimanfaatkan dan digunakan untuk usaha dan aktivitas transaksi bisnis dikenal dengan istilah *Electronic Commerce* (*e-commerce*).

Menurut McLeod (2008 : 59) perdagangan elektronik (e-Commerce) adalah penggunaan jaringan komunikasi dan komputer untuk melaksanakan proses bisnis [1]. Aktivitas transaksi bisnis yang dilakukan di e-commerce meliputi pemasaran, promosi, public relation, transaksi, pembayaran dan penjadwalan pengiriman barang, dan aktivitas lainnya tergantung dengan perkembangan teknologi dan keputusan dari penyedia layanan e-commerce [2]. Berdasarkan aktivitas transaksi tersebut akan diperoleh data yang berhubungan dengan e-commerce. Data tersebut dapat digunakan untuk mengelola dan memperoleh informasi yang berhubungan dengan product, customer, seller, dan lainnya. Namun jumlah data yang besar dan bertumbuh dari waktu ke waktu akan sulit diolah secara manual. Salah satu cara yang dapat dilakukan untuk mengelola dan memperoleh informasi dari data e-commerce tersebut adalah dengan visualisasi data. Berdasarkan buku "Visualization Design and Analysis: Abstractions, Principles, and Methods" yang ditulis oleh Tamara Muzner, definisi visualisasi adalah sistem visualisasi berbasis komputer yang memberikan representasi visual dari kumpulan data yang dimaksudkan untuk membantu pengguna (user) melaksanakan beberapa tugas dengan lebih baik. Sistem visualisasi juga dapat bersifat interaktif kepada pengguna (user). Terdapat juga

kemungkinan adanya keterbatasan sumber daya mencakup kapasitas komputer, manusia, dan *displays* [3].

Visualisasi data dapat dilakukan dengan menggunakan Tableau sebagai alat (tools) visualisasi yang diharapkan dapat mengelola data dengan baik. Data yang ditampilkan dengan visualisasi dapat lebih mudah untuk dipahami dan dianalisis oleh pengguna (user) dan lebih efektif untuk disajikan dalam bentuk visual daripada dalam bentuk angka. Visualisasi memungkinkan analisis untuk mengidentifikasi trend, menemukan pola, dan hal lain dari data yang digunakan. Dalam hal digunakan dataset dari Brazilian e-commerce. Dataset ini merupakan kumpulan data publik e-commerce Brazil yang dibuat Olist Store. Dataset Brazilian e-commerce memiliki informasi pesanan dari 2016 hingga 2018 yang diperolah dari di beberapa marketplaces di Brazil.

1.2 Tujuan

Tujuan dari penelitian ini adalah untuk memvisualisasikan *Brazilian E-Commerce Public Dataset by Olist* dengan menerapkan teknik-teknik dan *rule* dalam visualisasi sehingga dapat mengkomunikasikan informasi dan pengetahuan secara efisien melalui tampilan grafis.

1.3 Pertanyaan Visualisasi

Berikut pertanyaan penuntun visualisasi yang akan diimpelementasikan:

- 1. Bagaimana trend penjualan pada tahun 2016 sampai 2018 berdasarkan penjualan *product*? Negara apa tingkat penjualannya paling tinggi?
- 2. Wilayah mana yang memiliki jumlah *customer* paling banyak?
- 3. Wilayah mana yang memiliki jumlah seller yang paling banyak?
- 4. Berapa note rangking atau score review tertinggi yang dilakukan oleh customer?
- 5. *Product* manakah yang paling laris terjual disetiap bulan pada tahun 2018?
- 6. Berapakah hasil penjualan tertinggi mulai tahun 2016 sampai 2018?

1.4 Manfaat

Manfaat penelitian ini diharapkan dapat memberikan sumbangan pemikiran dalam memperkaya wawasan didunia akademik khususnya dalam penelitian yang dilaksanakan oleh peneliti di masa depan dalam hal visualisasi pada dataset *e-commerce*. Hasil visualisasi yang

diimplementasikan oleh tim diharapkan dapat memberikan informasi yang dapat membantu *e-commerce* mempermudah proses pengambilan keputusan.

1.5 Ruang Lingkup

Batasan dalam pengerjaan proyek ini adalah:

- 1. Data yang digunakan dalam visualisasi adalah B*razilian E-Commerce Public Dataset by Olist*
- 2. Dataset yang digunakan terdiri dari 9 file csv. yang tersedia di kaggle yaitu:
 - olist_customers_dataset.csv
 - olist_geolocation_dataset.csv
 - olist_order_items_dataset.csv
 - olist_order_payments_dataset.csv
 - olist_order_reviews_dataset.csv
 - olist_orders_dataset.csv
 - olist_products_dataset.csv
 - olist_sellers_dataset.csv
 - product_category_name_translation.csv
- 3. Tools yang digunakan untuk memvisualisasikan data adalah Tabelau

BAB 2 ISI

Bab ini akan menjelaskan pengertian visualisasi data, dan tahapan aktivitas dalam mengembangkan *dashboard* visualisasi data.

2.1 Visualisasi Data

Visualisasi data merupakan sebuah ilmu yang merepresentasikan penggambaran dari sebuah data dimana segala informasi telah melalui proses abstraksi ke dalam bentuk skematik, termasuk juga di dalamnya karakteristik ataupun variabel untuk setiap unit informasi. Tujuan diterapkannya visualisasi data adalah untuk mengkomunikasikan informasi maupun data secara jelas dan efisien dapat melalui penggambaran dengan menggunakan idiom-idiom visualisasi seperti grafik statistik, *scatter plot*, *bar chart*, *pie chart*, dan lain sebagainya sehingga memudahkan pembaca dalam memahami konsep dan dapat mengidentifikasi polapola yang terdapat dalam data. [5] Berikut merupakan beberapa manfaat yang didapat dengan menerapkan visualisasi data untuk kapasitas data yang besar:

a. Communicate Findings in Constructive Ways

Dengan adanya visualisasi data, maka dapat memungkinkan untuk merangkum keseluruhan informasi yang bersifat kompleks menjadi sebuah serangkaian gambaran sederhana.

b. Faster Action

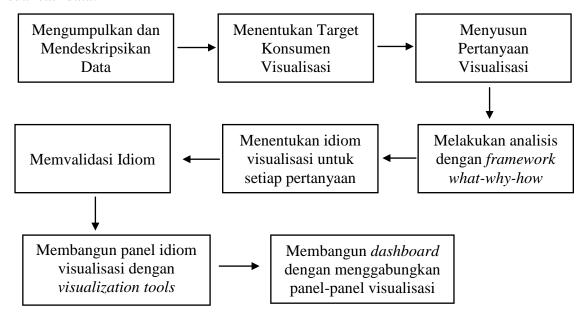
Memberikan bentuk komunikasi yang jelas dan memungkinkan beberapa pihak dalam melakukan penafsiran dalam pengambilan keputusan berdasarkan informasi yang didapat dari hasil visualisasi data dan juga dapat dengan lebih cepat memberikan respon yang sesuai terhadap perubahan kebutuhan pada kondisi tertentu.

c. Understand Connections Between Operations and Results

Visualisasi data memungkinkan pembaca dalam menemukan korelasi antara fungsi bisnis dan kinerja dalam lingkungan yang kompetitif. Maka dengan adanya visualisasi, pihak perusahaan ataupun organisasi dapat dengan tanggap menanggulangi apabila ditemukan kendala ataupun permasalahan.

2.2 Aktivitas Pengembangan Dashboard Visualisasi Data

Berikut adalah tahapan yang akan dilakukan oleh tim mahasiswa dalam mengembangkan visualisasi data:



Gambar 1 Tahapan Pengembangan Visualisasi Data

2.2.1 Mengumpulkan dan Mendeskripsikan Data

Data yang digunakan dalam visualisasi adalah Brazilian E-Commerce Public Dataset by Olist. Dataset yang digunakan terdiri dari 9 file csv. yang tersedia di kaggle yaitu:

- olist_customers_dataset.csv
- olist_geolocation_dataset.csv
- olist_order_items_dataset.csv
- olist_order_payments_dataset.csv
- olist_order_reviews_dataset.csv
- olist_orders_dataset.csv
- olist_products_dataset.csv
- olist_sellers_dataset.csv
- product_category_name_translation.csv

2.2.2 Menentukan Target Konsumen Visualisasi

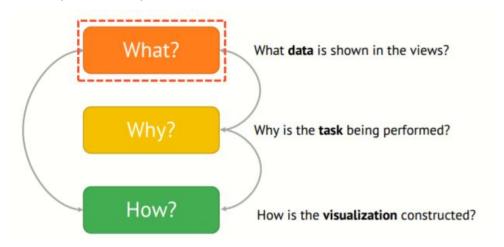
Target konsumen visualisasi adalah *e-commerce* dimana yang menjadi sampel dalam melakukan visualisasi adalah *Brizzillian E-Commerce*. Pengembangan visualisasi ini diharapkan dapat memberikan masukan dan sumbangan ilmu kepada berbagai *e-commerce* yang ingin memperoleh infomasi dari data mereka dengan teknik visualisasi data.

2.2.3 Menyusun Pertanyaan Visualisasi

Berikut pertanyaan penuntun visualisasi yang akan diimpelementasikan:

- 1. Bagaimana trend penjualan pada tahun 2016 sampai 2018 berdasarkan penjualan *product*? Negara apa tingkat penjualannya paling tinggi?
- 2. Wilayah mana yang memiliki jumlah *customer* paling banyak?
- 3. Wilayah mana yang memiliki jumlah *seller* yang paling banyak?
- 4. Berapa note *rangking* atau *score* review tertinggi yang dilakukan oleh *customer?*
- 5. *Product* manakah yang paling laris terjual disetiap bulan pada tahun 2018?
- 6. Berapakah hasil penjualan tertinggi mulai tahun 2016 sampai 2018?

2.2.4 What-Why-How Analysis Framework



Gambar 2 Kerangka kerja fFramework) menurut Tamara Muzner

Muzner pada tahun 2014 mengajukan sebuah kerangka kerja (*framework*) tingkat tinggi untuk menganalisis penggunaan visualisasi berdasarkan 3 (tiga) pertanyaan, yaitu:

- 1. What data is shown in the views?
- 2. Why is the **task** being performed?
- 3. How is the **visualization** constructed?

Ketiga pertanyaan tersebut memiliki tiga jawaban juga yaitu data, task dan *visualization* (idiom) yang sesuai. Setiap pertanyaan untuk kerangka kerja (*framework*) ini disebut sebagai *instance*.

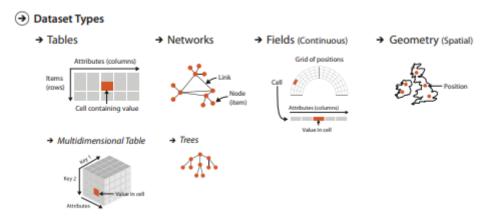
What data is shown in the views?

Pada tahapan ini, perlu didefinisikan dengan jelas jenis abstraksi data apa yang dapat divisualisasikan. Data adalah fakta atau statistik yang dikumpulkan bersama untuk referensi atau analysis. Data dapat berupa *images*, video, suara, text, dan *log files*. Banyak aspek

visualisasi yang didorong dari jenis data yang dimiliki. Data bisa saja memiliki interpretasi yang berbeda-beda, maka dari itu perlu diketahui *semantic* data atau makna sebenarnya didunia nyata.

Data yang berada dalam satu kumpulan atau disebut dataset terdiri atas 4 tipe dasar yaitu:

- Tables terdiri atas rows yang merepresentasikan item data dan columns yang merepresentasikan attribute dataset. Setiap cell dalam table merupakan spesifikasi dari kombinasi row dan column.
- *Networks* terdiri atas item yang disebut *node* dan *link* yang merupakan penghubung antara 2 *items*.
- Fields terdiri dari nilai yang berasosiasi dengan *cell*. Setiap *cell* mengandung ukuran atau kalkulasi dari domain *continuous*.
- *Geometry*. jenis kumpulan data geometri menentukan informasi tentang bentuk item dan posisi spasial eksplisit. Item dapat berupa titik, garis atau kurva satu dimensi, bidang dua dimensi atau volume 3 dimensi.



Gambar 3 Tipe Dataset menurut Tamara Munzner

Dataset dibentuk dari kombinasi beberapa tipe data atau disebut *data types*. Terdapat 5 *data types* yaitu *items, attributes, links, positions,* dan *grids*.



Gambar 4 Tipe Data menurut Tamara Munzner

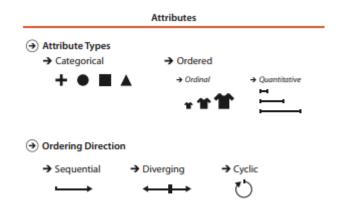
Berikut adalah kombinasi antara tipe dataset dengan data menurut Tamara Munzner.



Gambar 5 Data dan Dataset Types menurut Tamara Munzner

Selanjutnya adalah attribute pada dataset yang terdiri atas dua tipe data secara umum yaitu :

- *Categorical* (nominal) nilai pada atribut memiliki bobot yang sama dan menggunakan istilah yang digunakan sebagai lambang. Nilainya tidak dapat dilakukan perhitungan matematika. Contoh: Nama, *gender*, dan ID.
- Ordered: Ordinal dan Quantitative
 - *Ordinal*: nilai memiliki relasi lebih besar atau lebih kecil sehingga terdapat urutan atau *order* namun pada nilai tidak dapat dilakukan perhitungan matematika. Contoh: hari dalam seminggu, *grade*, bulan, dan lain sebagainya.
 - *Quantitative*, terdiri dari 2 tipe yaitu *Interval* dimana nilai dapat dilakukan perhitungan matematika namun nilai 0 adalah *arbitary*. Contoh : tanggal. Tipe kedua adalah *Ratio* dimana dapat nilai dapat dilakukan perhitungan matematikan dan *fixed zero*. Contoh : *temperature*, jumlah stok, dan lain sebagainya.

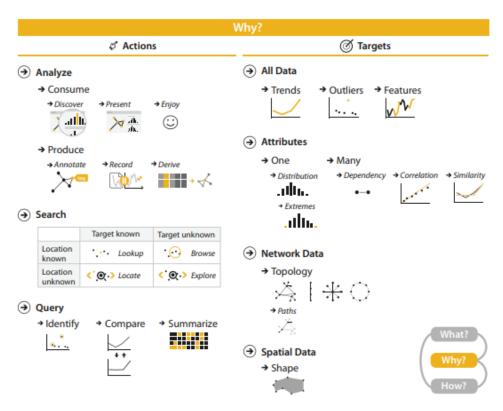


Gambar 6 Attribute Types menurut Tamara Munzner

Why is the task being performed?

Pada tahapan ini perlu didefinisikan *task abstraction* dengan memecah menjadi *actions* dan *targets*. Menurut Tamara Munzner Level tertinggi dari *actions* adalah menggunakan visualisasi untuk mengkonsumsi/*consume* atau menghasilkan/*produce* informasi. Kasuskasus pada *consume* adalah untuk menyajikan/*present*, menemukan/*discover*,

menikmati/enjoy. Pada Level menengah, visualisasi dapat digunakan untuk melakukan pencarian/search berdasarkan indentitas dan location target diketahui atau tidak. Pada level rendah, terdapat query yang memiliki 3 cakupan yaitu : identify one target, compare some target, and summarize all targets. Framework ini mendorong untuk mempertimbangkan tugas dalam bentuk abstrak. Actions yang sudah dibahas sebelumnya mengacu pada target, yang berarti setiap aspek data menarik bagi pengguna. Target adalah nouns sementara actions adalah verbs.



Gambar 7 Actions dan Target menurut Munzner

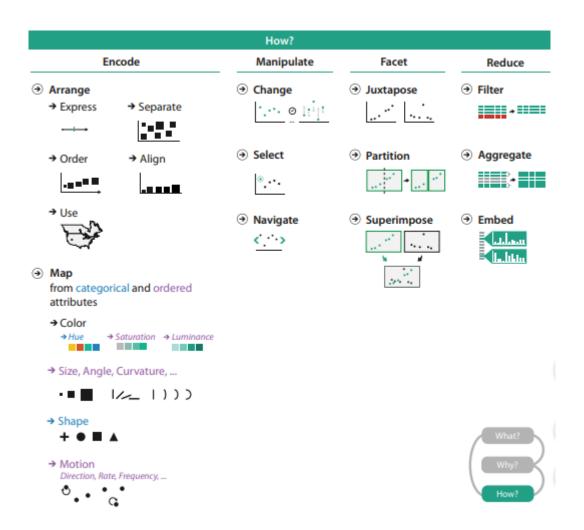
How is the visualization constructed?

Tahapan selanjutnya adalah bagaimana membangun idiom visualisasi sesuai dengan *data and task abstraction*. Kemudian bagaimana idiom visualisasi dapat dibangun dari serangkaian pilihan desain oleh Tamara Munzner yang dapat ditinjau dari 4 kelas utama yaitu :

• Encode

- *Arrange* terdapat 5 cara untuk menyusun data secara spasial yaitu *express values*, *separate*, *order*, *align regions*, dan *use*
- *Map* mencakup bagaimana memetakan data dengan semua visual *channels* nonspasial yaitu *color*, *size*, *angle*, *shape*, dsb.

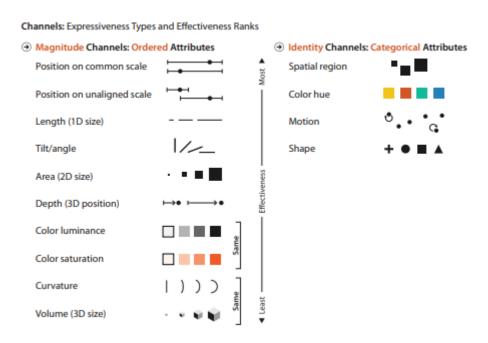
- Manipulate memiliki pilihan untuk mengubah setiap aspek tampilan (change), memilih elemen dari dalam tampilan (select), dan menavigasi untuk mengubah sudut pandang dalam tampilan (navigate)
- Facet yaitu bagaimana melakukan juxtapose (menyandingkan) dan mengoordinasikan beberapa tampilan, bagaimana mempartisi data diantara tampilan, dan bagaimana menempatkan lapisan diatas satu sama lain (superimpose)
- Reduce yaitu bagaimana cara mengurangi (*reduce*) data yang ditampilkan dengan memiliki opsi untuk menyaring data (*filter*), menggabungkan banyak elemen menjadi satu (*aggregate*, dan menyematkan informasi utama dalam satu tampilan (*embed*).



Gambar 8 How to design vis idiom: encode, manipulate, facet, and reduce by Tamara Munzner

Marks and Channel

Marks adalah elemen geometris dasar yang menggambarkan item atau links, dan channel akan mengontrol tampilannya. Efektivitas channel untuk melakukan encoding pada data bergantung pada tipe datanya. Tim perlu menentukan tipe marks and channel yang akan digunakan pada idiom visualisasi yang sudah ditentukan. Pemilihan marks and channel menggunakan peringkat efektifitas channel dan tipe-tipe mark visualisasi menurut Tamara Munzner.



Gambar 9 Expressiveness Types and Effectiveness Ranks by Tamara Munzner

2.2.5 Validation

Tahap selanjutnya adalah melakukan validasi terhadap visualisasi yang sudah dirancang melalui 4 level desain yaitu domain situation, task and data abstraction, visual encoding and interactio, dan algorithm. Dimana level task and data abstraction membahas pertanyaan why dan what, dan level idiom membahas pertanyaan how yang sebelumnya sudah dibahas. Setiap lebel memiliki threats yang berbeda-beda, maka dari itu perlu memilih metode validasi yang baik dan sesuai. Validasi perlu dilakukan karena sebagian besar desain visualisasi tidak efektif dalam menyampaikan informasi.

Tipe evaluasi yang dilakukan adalah dengan Quantitative secara Formative yaitu sebelum atau selama pengembangan karena lebih mudah untuk membandingkan hasil dan dapat diukur dengan mudah. Mekanisme evaluasi menurut Munzner's Nested Model:



1. Domain

Threat: wrong problem

Validate: melakukan observasi dan interview pada target user

2. Data/task abstraction

Threat: wrong task/data abstraction

Validate: testing/pengujian pada target *user*, studi lapangan, *collect* infomasi mengenai data (komponen "**What"dan "Why"**)

3. Idiom

Threat: Ineffective encoding/interaction technique

Validate :Menerapkan komponen "How" yaitu mengevaluasi bagaimana visual encoding, Qualitative/quantitative result image analysis, Lab study, measure human time/errors for task Validate Justify encoding/interaction design Implement system Test on any users, informal usability study

4. Algorithm

Threat: slow algorithm

Validate: menganalisis kompleksitas komputasi, mengukur time/memori sistem.

2.2.6 Membangun Panel Idiom dan Dashboard Visualisasi dengan Visualization Tool

Setelah menentukan dan menganalisis idiom visualisasi yang akan dibangun, selanjutnya adalah membangun panel visualisasi sesuai dengan pertanyaan dan kemudian menggabung panel-panel dalam sebuah dashboard. Dalam proses ini, tim menggunakan Tableau visualization tool.

BAB 3 RENCANA KERJA

3.1 Jadwal Kegiatan

Pengerjaan proyek dimulai pada minggu ke 11 yaitu pada tanggal 17 November 2020. Tim melakukan *brainstorming* dan menentukan topik yang akan dibahas pada proyek di minggu pertama. Tim juga menyusun proposal proyek pada minggu ke 1, proposal proyek terdiri dari 3 bab yaitu Pendahuluan, Landasan Teori, dan Rencana Kerja.Proposal dikumpulkan pada Minggu, 22 November 2020 yaitu di minggu ke 2. Setelah itu tim mulai mengerjakan visualisasi data dan membuat *dashboard* sesuai dengan proposal yang diajukan di minggu ke 2 dan 3. Penyusunan laporan akhir dilakukan pada minggu ke 2, 3 dan 4 sesuai dengan implementasi yang dilakukan. Pada minggu ke 4 tim juga membuat video presentasi sesuai dengan implementasi dan laporan akhir. Laporan akhir dan video presentasi dikumpulkan di minggu ke 4 pada Jumat, 8 Januari 2021. Pengerjaan proyek dimulai pada 17 November 2020 dan dikumpulkan pada 8 Januari 2021, jadi total waktu (dalam minggu) pengerjaan proyek adalah 4 minggu.u. Jadwal Kegiatan pengerjaan proyek dapat dilihat pada Tabel 2 berikut ini.

Tabel 1 Jadwal Kegiatan

Kegiatan		Minggu			
	1	2	3	4	
Brainstorming dan Menentukan Topik					
Menyusun Proposal					
Pengumpulan Proposal					
Mengerjakan Dashboard Visualisasi (Implementasi)					
Menyusun Laporan Akhir					
Membuat Video Presentasi					
Pengumpulan Laporan Akhir dan Membuat Video Presentasi					

3.2 Team Member and Roles

Berikut merupakan tabel pembagian tugas setiap anggota kelompok.

Tabel 2 Team Members and Tasks

Team Members	Tasks			
Heppy Maria Simanungkalit	 Menyusun Proposal Membuat panel visualisasi untuk pertanyaan 1 dan 2 Menyusun Laporan Akhir Membuat slide dan video presentasi 			
Evola R A Tampubolon	 Menyusun Proposal Membuat panel visualisasi untuk pertanyaan 3 dan 4 Menyusun Laporan Akhir Membuat slide dan video presentasi 			
Tri Dessy Natalia	 Menyusun Proposal Membuat panel visualisasi untuk pertanyaan 5 dan 6 Menyusun Laporan Akhir Membuat slide dan video presentasi 			

Daftar Pustaka

- [1] S. M. Maulana, H. Susilo dan Riyadi, "Implementasi E-Commerce Sebagai Media Penjualan Online (Studi Kasus Pada Toko Pastbrik Kota Malang)," *Jurnal Administrasi Bisnis (JAB)*, vol. 29, no. 1, pp. 1-9, 2015.
- [2] S. Avriyanti, "Peran E-Commerce Untuk Meningkatkan Keunggulan Kompetitif Di Era Industri 4.0 (Studi Pada Ukm Yang Terdaftar Pada Dinas Koperasi, Usaha Kecil Dan Menengah Kabupaten Tabalong)," *Jurnal PubBis*, vol. IV, no. 1, pp. 82-99, 2020.
- [3] T. Munzner, Visualization Design and Analysis: Abstractions, Principles, and Methods, Boca Raton: A K Peters/CRC Press, 2012.
- [4] T. Munzner, Visualization Analysis and Design, Boca Raton: A K Peters/CRC Press, 2014.
- [5]I. R. Widyan, "PLATFORM VISUALISASI DATA UNTUK PEMERINTAH AMSTERDAM SEBAGAI SOLUSI PEMBERSIHAN KOTA SECARA EFEKTIF," Surabaya, 2017.