Part 1. Analytics [30 marks total]

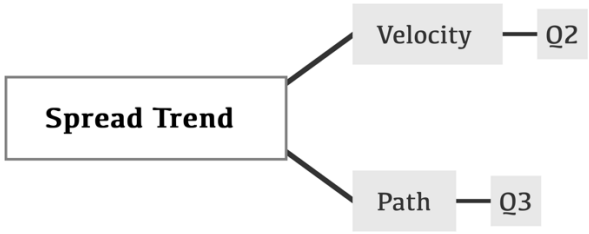
a.

Q1: Analyze the spread trend of this virus all over the world. What is the spread over time?

Q2: What is the monthly death rate of coronavirus in each city?

Q3: What date does the first case happen?

*Mind map about exploring Q2 & Q3



b.

	File	Items	Attributes
(required)	COVID-19-geographic-disbtribution-worldwide.csv	10947 items, including 206 countries during 2019/12/31 to 2020/04/15	dateRep: the date has cases day: day of dateRep month: month of dateRep year: year of dateRep cases: number of confirmed cases each day deaths: number of deaths each day countriesAndTerritories: country name geoId: abbreviation of country name countryterritoryCode: abbreviation of territory name popData2018: population in 2018 in the same month

(others)	WHO-COVID-19-global-data.csv	12879 items, including 215 countries during 2020/1/11 to 2020/5/10	day: the date has cases Country: abbreviation of country name Country Name: country full name Region: continent that country belongs to Deaths: number of deaths each day Cumulative Deaths: number of deaths until the date Confirmed: number of confirmed cases each day Cumulative Confirmed: number of cases until the date

c.

File	Attributes	Attribute Types	Attribute Semantics		
			Spatial	Temporal	C/D
COVID-19-geographic-disbtribution-worldwide.csv	dateRep	quantitative	×	✓	Discrete
	day	hierarchical	×	✓	-
	month	hierarchical	×	✓	-
	year	hierarchical	×	✓	-
	cases	quantitative	×	×	Continuous
	deaths	quantitative	×	×	Continuous
	countriesAndTerritories	categorical	×	×	-
	geoId	categorical	×	×	-
	countryterritoryCode	categorical	×	×	-
	popData2018	quantitative	×	×	Continuous
WHO-COVID-19-global-data.csv	day	quantitative	×	✓	Discrete
	Country	categorical	×	×	-
	Country Name	categorical	×	×	-
	Region	categorical	×	×	-
	Deaths	quantitative	×	×	Continuous
	Cumulative Deaths	quantitative	×	×	Continuous
	Confirmed	quantitative	×	×	Continuous
	Cumulative Confirmed	quantitative	×	×	Continuous

Correlation:

1. In "WHO-COVID-19-global-data.csv ", "cumulative deaths/confirmed" and "deaths/confirmed" is the cumulative deaths/confirmed in day (n-1) + deaths/confirmed in day n = cumulative deaths/confirmed in day n.

2. In “COVID-19-geographic-disbtribution-worldwide.csv”, “day”, “month” and “year” is sub-datas of “date”.

3.Between two files, they have similar country name and have some same time period.

link of “WHO-COVID-19-global-data.csv”