**Topic 1**

Business Problem : *Do exit polls influence voting behavior / election results*

Why do you think it’s important ?

1. Elections decide country’s future for 5 years. Unbiased and uninfluenced results would bring us the people’s chosen winner.
2. But what if the media, to increase their TRPs, is going overboard with their exit polls in bombarding people with their biased or influencing predictions swinging the results.

Data sources : NY times Exit polls data sets for each state and each district

Methods : Compare actual results with these exit polls, train model and validate it. Create surveys for demographics of samples from each state (test data) to see who will be potential winner in that area in 2020

**Topic 2 :**

Business Problem :

1. *Predicting the price of a used car based on car and location attributes*
2. *Car selection using personality traits of individuals*

Why do you think it’s important ? :

* The percentage of middle class people is approximately 50% and the number of students who tend to buy used cars is also significantly high.
* The market for used cars is huge and choosing the right car that fits the budget isn’t always easy given the options.
* Buyer behaviors affect decision making, it would be interesting to see what personality traits of people would lead them to buying cars.

Data sources :

* Data available on kaggle.com, 1.3 M records
* Variables such as manufacture, make, year, cylinders, latitude, longitude, color, description etc.
* Web scraping of user reviews on different cars from kbb.com

(<https://www.kbb.com/car-reviews/>)

* A custom designed questionnaire survey for people to assess personality traits

Consider : Supply and demand. How do we account this in data ?

**Topic 3**

Business Problem : Predict the fast submerging cities/areas in the world due to climate change/ growing sea levels.

Why do you think it’s important ?

Climate change is the biggest problem right now. With recent Amazon fires, any insights on these lines would draw attention. Many major cities in the world are on coastal areas of their respective countries. Jakarta’s predicted to be under water in next 10 years. It’s important to understand the gravity of the situation and start curbing mechanism if we should.

Global avg. sea level is increasing by 1 inch/decade. Rising seas is [one of those climate change effects](https://www.nationalgeographic.com/environment/global-warming/global-warming-effects/). Average sea levels have swelled [over 8 inches (about 23 cm) since 1880, with about three of those inches gained in the last 25 years.](https://www.globalchange.gov/browse/indicators/global-sea-level-rise) Every year, the sea [rises another .13 inches (3.2 mm)](https://climate.nasa.gov/vital-signs/sea-level/).

Data sources : International / National Meteorological depts websites - Inputs could be : Altitudes of city above/below sea level, local rate of increase in sea level, any infrastructure variables that imply bad planning/population explosion/sewer system planning (does it matter?), any government plans that are working towards saving the city from the calamity etc.

Potential causes :

1)Sea water temperature rise

2) Melting of ice on land adds more water

3)Illegal wells to extract groundwater

4)Extraction of oil and gas

4)Weight of buildings due to urbanization

5)Coastal development rate (Cutting down of mangroves/forests, they are natural flood barriers.)

6)Warm air, warm water temperatures

7)Altitude above sea level

8)Rain fall

9)VEgetation

10) Ice cover

(Illegal wells digged to access groundwater. Because more than 97 percent of the city is covered in concrete, the groundwater is not replenished by rain and rivers. The city is also sinking due to the weight of its buildings. In addition, natural flood barriers like mangroves have been cut down to clear space for housing.)

Data source <https://tidesandcurrents.noaa.gov/inundation/>

**Topic 4**

Business Problem :

*To predict whether Bitcoin transactions belong to licit categories (exchanges, wallet providers, miners, licit services, etc.) versus illicit ones (scams, malware, terrorist organizations, ransomware, Ponzi schemes, etc.)*

Why do you think it’s important ? :

* When released in 2009, the digital currency had a value of less than one U.S. penny per “coin.” Now, just eight years later, one bitcoin is worth over $15,000 and the cryptocurrency’s market cap is over $200 billion.
* Recent research by Chainanalysis suggests that the amount of bitcoin spent on illegal transactions this year could hit a record high of $1 billion
* **As the U.S. continues to see exponential growth in the consumption of medical marijana (cannabis), the product is still considered illegal at the Federal level and thus restricts Federally insured Banks from partnering with cannabis organizations. The use of bitcoin or other cryptocurrency in the cannabis marketplace could mitigate the stigma of illicit use to one of legitimate business means. Thereby elevating both cryptocurrency and cannabis use in mainstream America.**

Data sources :

* Data available on kaggle.com

<https://www.kaggle.com/ellipticco/elliptic-data-set>

**Topic 5**

Business Problem :

*To find patterns from 900k hands of Black Jack and find out a strategy to optimize wins*

Why do you think it’s important ? :

* Blackjack is undoubtedly one of the most popular table games in any casino, fairly simple with few rules and a straightforward two-fold premise
* Reasons why most people lose are not checking the rules of the game, trying to beat the house with a betting, making the insurance bet and so on
* Through a strategy that optimizes winning chances, users can enjoy the game and make money out of it.
* **Another reason it might be important to identify patterns in Black Jack is to test if the gaming establishments are playing by the rules of random chance or applying a probability algorithm that improves the odds for the house.**

Data sources :

* Data available on kaggle.com
* <https://www.kaggle.com/mojocolors/900000-hands-of-blackjack-results>

**Topic 6**

Business Problem :

*Quality prediction of how much impurity is in the ore concentrate in a mining process.*

Why do you think it’s important ? :

* The most important part of a mining process is a [flotation plant](https://en.wikipedia.org/wiki/Froth_flotation)! this impurity is measured every hour, if we can predict how much silica (impurity) is in the ore concentrate, we can help the engineers, giving them early information to take actions (empowering!). Hence, they will be able to take corrective actions in advance (reduce impurity, if it is the case) and also help the environment (reducing the amount of ore that goes to tailings as you reduce silica in the ore concentrate).

Data sources :

* Data available on kaggle.com

<https://www.kaggle.com/edumagalhaes/quality-prediction-in-a-mining-process>

**Topic 7**

Business Problem :

*Deciphering what drives the demand of uber and lyft ride prices and also discover people’s preferences of one over the other*

Why do you think it’s important ? :

* Uber and Lyft's ride prices are not constant like public transport. They are greatly affected by the demand and supply of rides at a given time.
* time of the day for example; times around 9 am and 5 pm should see the highest surges on account of people commuting to work/home. Another guess would be the weather; rain/snow should cause more people to take rides.
* **Another possible predictor of what drives demand for uber and lyft prices is the driver’s revenue share with either uber and lyft. As there is no exclusivity of which service a driver participates with, there may also be a driver bias that influences customer preferences.**

Data sources :

* Data available on kaggle.com
* <https://www.kaggle.com/ravi72munde/uber-lyft-cab-prices>

**Topic 8**

Business Problem :

*Predict and Assess the forest fires in Brazil*

Why do you think it’s important ? :

* Forest fires are a serious problem for the preservation of the Tropical Forests. Understanding the frequency of forest fires in a time series can help to take action to prevent them. Brazil has the largest rainforest on the planet that is the Amazon rainforest.
* With this data, it is possible to assess the evolution of fires over the years as well as the regions where they were concentrated.

Data sources :

* Data available on kaggle.com

<https://www.kaggle.com/gustavomodelli/forest-fires-in-brazil>

**Topic 9**

Business Problem :

*Predict and Assess the impact of U.S. and China trade war as well as the effect of Brexit on North American Trade.*

Why do you think it is important?

* The current U.S. administration is reevaluating its trade agreements with China and North America to the unilateral benefit of the U.S.. In response, its largest trading partner, Canada is likely to pivot its trading with the U.S. and thus manifesting in less border crossings and economic prosperity. A predictive model utilizing different trade scenarios could be applied to determine the economic impact to U.S. border towns.

Ideas with datasets but no business problem-

1. **Border Crossing Entry Dataset**- The Bureau of Transportation Statistics (BTS) Border Crossing Data provide summary statistics for inbound crossings at the U.S.-Canada and the U.S.-Mexico border at the port level. Data are available for trucks, trains, containers, buses, personal vehicles, passengers, and pedestrians. Border crossing data are collected at ports of entry by U.S. Customs and Border Protection (CBP). The data reflect the number of vehicles, containers, passengers or pedestrians entering the United States.

Link to the dataset-<https://www.kaggle.com/akhilv11/border-crossing-entry-data>

1. **NBA Players dataset-** The data set contains two decades of data on each player who has been part of a team's roster. It captures demographic variables such as age, height, weight and place of birth, biographical details like the team played for, draft year and round. In addition, it has basic box score statistics such as games played, average number of points, rebounds, assists, etc.

Analysis Ideas

The data set can be used to explore how age/height/weight tendencies have changed over time due to changes in game philosophy and player development strategies. Also, it could be interesting to see how geographically diverse the NBA is and how oversees talents have influenced it. A longitudinal study on players' career arches can also be performed.

<https://www.kaggle.com/justinas/nba-players-data>

Ideas with business problem but no dataset:

1. Predict the salary of a person based on resume
2. Extracting summaries from articles to reduce the burden of going through the entire article
3. Predict customers who are more likely to return items purchased from amazon or walmart within the 30 day period