# **MODELING FUTURE CONTRACEPTIVE USE**

1. **BUSINESS UNDERSTANDING**

***Overview***

Contraceptive utilization is the practice that helps individuals or couples avoid unwanted pregnancy. Beyond preventing unwanted pregnancy, contraceptive access is vital to safe motherhood and prosperous communities.

Access to contraceptives enables individuals/couples to determine when they want to begin having children, how far apart they want their children to be and when to stop having children. It also slows unsustainable population growth. This is important because overpopulation puts pressure on the economy affecting services being offered to citizens.

[Contraceptive methods](https://www.cdc.gov/reproductivehealth/contraception/index.htm) range from The condom, The oral contraceptive pill, Intrauterine Device (IUD), The Contraceptive Implant, The Contraceptive Injection, Emergency Contraception Pill (The ‘Morning After’ Pill), Contraceptive Ring, Diaphragm to Sterilization.

It is important that contraceptives are widely available and easily accessible to anyone who is sexually active. According to the [UN Contraceptive use by method 2019 Data Booklet](https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/files/documents/2020/Jan/un_2019_contraceptiveusebymethod_databooklet.pdf) Africa has by far the lowest percentage of women using contraceptives. To be specific, [214 million](https://www.who.int/health-topics/contraception#tab=tab_1) women of reproductive age in developing regions have an unmet need for contraception.

One of the main reasons for limited access is that health systems in these countries are unable to predict the quantity of contraceptives necessary for each health service delivery site mainly due to inadequate systems and insufficient data. When too few supplies are ordered, service delivery sites run out limiting access to contraceptives and family planning. When too much product is ordered, unused contraceptives are wasted and left to expire.

Accurate forecasting of contraceptive consumption can save lives, money, and time by ensuring health service delivery sites have what they need when they need it and by reducing waste in the supply chain.

***Objectives***

[USAID](https://www.usaid.gov/global-health/health-systems-innovation) works with local health care authorities and partners to support voluntary family planning and reproductive health programs in nearly 40 countries across the globe. This includes ensuring contraceptives are available and accessible to people who need them.

West African countries are poised for significant change in family planning behaviors.

Particularly, findings from the [Family planning demand analysis](https://www.psi.org/wp-content/uploads/2020/02/Transform-CI-Segmentation_Final-Report_EN_vF.pdf) highlighted that based on the unmet demand for Family Planning (FP) alone, there are more than 2 million potential new contraceptive users in Côte d’Ivoire.

The **Main** goal of this project is to therefore build a model that predicts consumption for 11 contraceptives across 156 health service delivery sites in the public sector health system in Côte d’Ivoire. The predictions are to be made monthly from October to December 2019.

The **Specific** objectives include:

1. Identifying the most popular contraceptive product .
2. Identifying the most visited site so as to
3. Finding out the most preferred form of contraceptive method.
4. Identifying stock distributed as per the different regions.
5. Finding the stock distributed by region and product code.

**2. DATA UNDERSTANDING**

***Initial Data***

The primary dataset(Train.csv) is extracted from the [Electronic Logistics Management Information System](https://openlmis.org/implementation_region/cote-divoire/) (eLMIS) used in Côte d’Ivoire to manage contraceptives inventory and distribution data at health service delivery site level.

The Secondary datasets( monthly\_cases. Csv, and annual\_cases.csv ) are extracted from the [District Health Information Software 2 Tracker](https://innov.afro.who.int/emerging-technological-innovations/dhis2-tracker-2041) (DHIS2) system that captures routine annual(2016-2018) and monthly(Jan-Sep 2019) facility data pertaining to contraceptive use aggregated at the district level. It is likely that there are sites reporting to this system in each district that do not report to the eLMIS system (the primary data source). There may also be discrepancies between number of districts and district names in these datasets compared to the primary dataset.

***Data Description***

The files available for download were:

* **train.csv**: the primary dataset that includes the target variable(stock\_distributed)
* **monthly\_cases.csv**: monthly information about distribution of contraceptives from January to June 2019
* **annual\_cases.csv**: annual information about distribution of contraceptives per health center for 3 years
* **samplesubmission.csv**: an example of what the submission file should look like
* **data\_dictionary.csv**: variable definition for each csv file
* **product.csv**: definition of each contraceptive product(Product Code)
* **recommended\_supplementary\_data\_sources.docx**: suggested publicly available external datasets that may be used as recommended by USAID
* **service\_delivery\_site\_data.csv**: information about health service delivery sites(Site Code) in the train data
* **Table\_of\_contents.xlsx**: description of the files available
* **contraceptive\_case\_data\_annual.csv**: annual information about distribution of contraceptives per health center for 3 years.

Description of the variables in the primary dataset are as follows:

* **Year**: year of the observation
* **Month**: month of the observation
* **Region:** higher level geographical area within Côte d'Ivoire
* **District:** lower level geographical area within Côte d'Ivoire
* **Site\_code:** actual health service delivery site
* **Product\_code:** unique identifier for each commodity
* **Stock\_initial:** stock in hand of previous period. This is quantified in dispensing units
* **Stock\_received:** total quantity received in the last period. This is quantified in dispensing units
* **Stock\_distributed{Target Variable}**: quantity dispensed/consumed in the last reporting period. This is quantified in dispensing units
* **Stock\_adjustment:** all kinds of losses/ adjustments made at the facility
* **Stock\_end:** current physical count of stock in hand. This is quantified in dispensing units
* **Average\_monthly\_consumption**: average monthly consumption for the last three months. This is quantified in dispensing units
* **Stock\_stockout\_days:** total number of days facility was out of stock
* **Stock\_ordered:** the requested quantity. This is quantified in dispensing units.

***Data Exploration***

The Primary dataset had a total of 35,753 columns and 14 rows. A total of 763 observations were observed as missing on the Stock Ordered column.

***Data Verification***

Based on the objectives of the Project, the data was fit to implement the goals.

Under data quality, not all sites report all data each month. There were also cases in the Primary dataset where the eLMIS system records a zero value that may in fact represent a non-reported value or a value entered by a user for convenience. In such instances, these zeros did not reflect actual observed values.

**3. DATA PREPARATION**

This step involved preparing the data for modeling.

***Data Selection***

From the datasets provided, the Primary dataset was used for modeling. The dataset was merged with the datasets providing information on delivery sites and types of contraceptives so as to provide proper description of the Site and Product codes.

The total number of columns were 19 after merging, namely: Year, Month, Region, District, Site code, Product code, Stock initial, Stock received, Stock distributed, Stock adjustment, Stock end, Average monthly consumption, Stock stockout days, Stock ordered, Site type, Site latitude, Site longitude and Product type.

***Data Cleaning***

* Missing Values

The missing values in the datasets were as follows:

Primary Dataset:

stock\_ordered 763

Since the dataset had a total of 35,753 rows, the missing entries were dropped as they would not interfere with the analysis.

Secondary Datasets:

Monthly cases:

Injection2 stock end 678

Iud stock end 678

Iud number dispensed 429

Iud women old 428

Iud women new 372

Iud number received 344

Injection2 women new 197

Injection2 number received 182

Injection2 number dispensed 170

Injection2 women old 122

Pill women new 83

Implant women old 68

Pill number dispensed 63

Pill number received 63

Pill women old 57

Pill stock end 52

Implant number received 35

Implant stock end 33

Implant number dispensed 29

Injection3 number received 21

Injection3 stock end 19

Injection3 number dispensed 16

Implant women new 6

Annual cases:

Injection2 stock end 339

Iud stock end 339

Iud women old 127

Iud number dispensed 106

Iud number received 94

Pill stock end 73

Implant stock end 73

Injection3 stock end 68

Iud women new 62

Implant women old 13

Injection2 number received 11

Injection2 number dispensed 8

Pill number received 7

Implant number received 6

Implant number dispensed 6

Pill number dispensed 5

Injection2 women new 5

Injection3 number received 2

Injection2 women old 2

Implant women new 2

Injection3 number dispensed 1

Pill women new 1

Due to the many missing entries in the Secondary datasets, they were not used for any further analysis.

* Duplicate Values

The Primary dataset lacked identical entries.

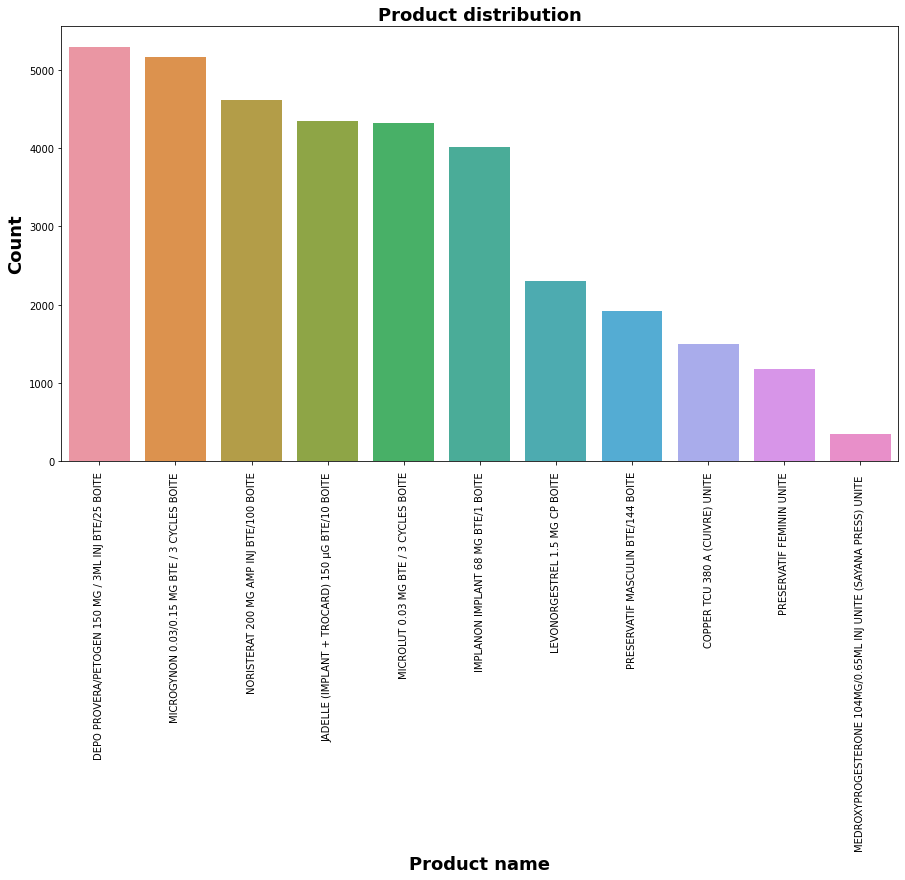
* Outliers

Outliers could be observed in all the columns in the dataset. Since there was no basis of assuming the entries were all not valid, no outliers were dropped. Dropping them would also mean losing most of the information needed for modeling.

**4. DATA ANALYSIS**

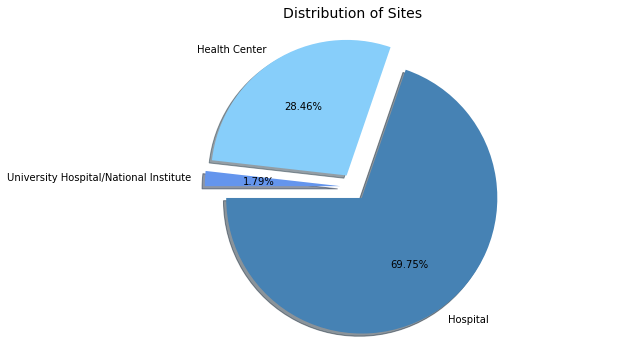
This process involved investigating the dataset to discover patterns. The analysis seeked to provide insight on:

1. Product Distribution

****

The most popular contraceptive was Injectable Contraceptive(DEPO PROVERA 150mg) followed by Oral Contraceptive Pill(MICROGYNON) while the The least popular contraceptive is Injectable Contraceptive(MEDROXYPROGESTERONE 104mg)

1. Site Distribution



According to the [Côte d’Ivoire Private Health Sector Assessment](https://shopsplusproject.org/sites/default/files/resources/C%C3%B4te%20d%E2%80%99Ivoire%20Private%20Health%20Sector%20Assessment_%20Family%20Planning_0.pdf), the provision of care is dominated by the public sector. These public health facilities are an important source of family planning services in Côte d’Ivoire accounting for 61% of the supply of modern family planning methods.

There are three levels of service delivery facilities in the public sector:

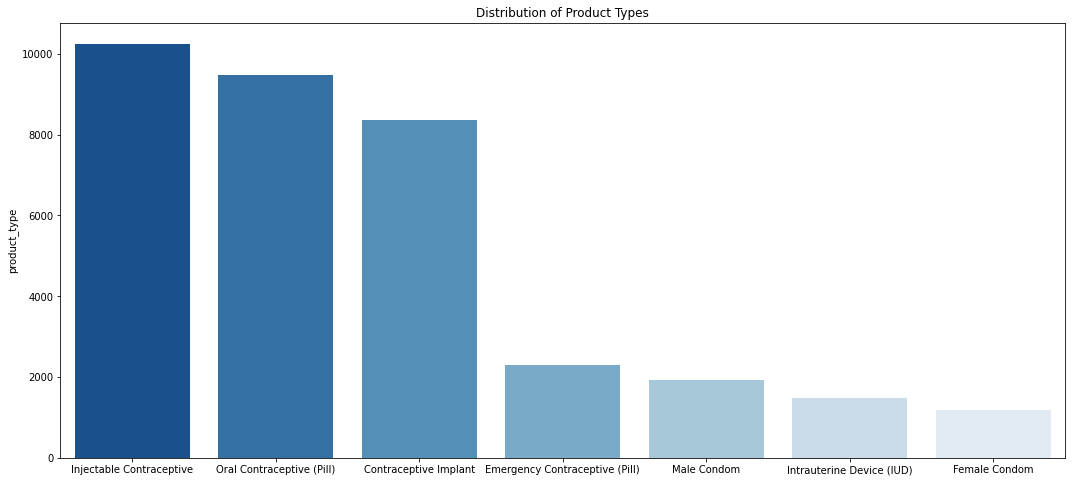
- Level 1: Public primary health care institutions

- Level 2: General,Regional, and Specialized hospitals at the secondary level

- Level 3: University teaching hospitals and Specialized health institutes at the Tertiary Level

The majority of health facilities in the public sector are categorized as Level 1 and Level 2 which comprises the General Hospitals.

1. Product Type Distribution



In Côte d’Ivoire, there is heavy reliance on Donor financing and Donations for Family planning products.

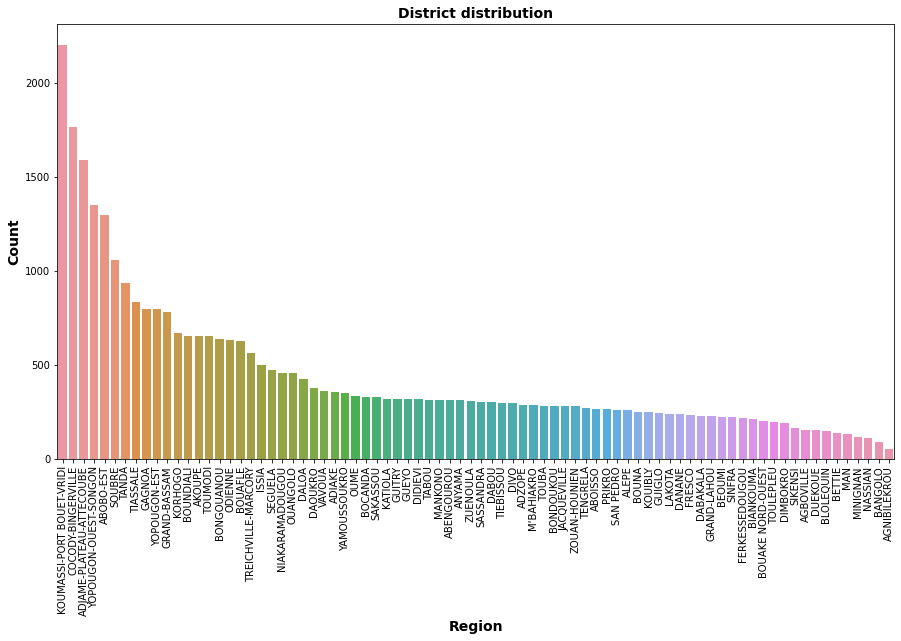
Referencing the [Côte d’Ivoire Private Health Sector Assessment](https://shopsplusproject.org/sites/default/files/resources/C%C3%B4te%20d%E2%80%99Ivoire%20Private%20Health%20Sector%20Assessment_%20Family%20Planning_0.pdf), the proportional value of donated contraceptive methods as of 2017 was as follows:

* 32% ($3.6M) for Male and Female Condoms
* 27% ($3.1M) for Injectables
* 21% ($2.4M) for Implants
* 18% ($2.0M) for Oral Pills
* 2% ($0.2M) for Emergency Contraceptives
* less than 1% for IUDs and other methods.

Most of these products are channeled through the public sector but some are also distributed through NGOs.

| **Product Type** | **Number of Brands** |
| --- | --- |
| Oral Pills | 19+ |
| Emergency Contraceptive | 9+ |
| Injectables | 1 |
| Condoms | 17+ |

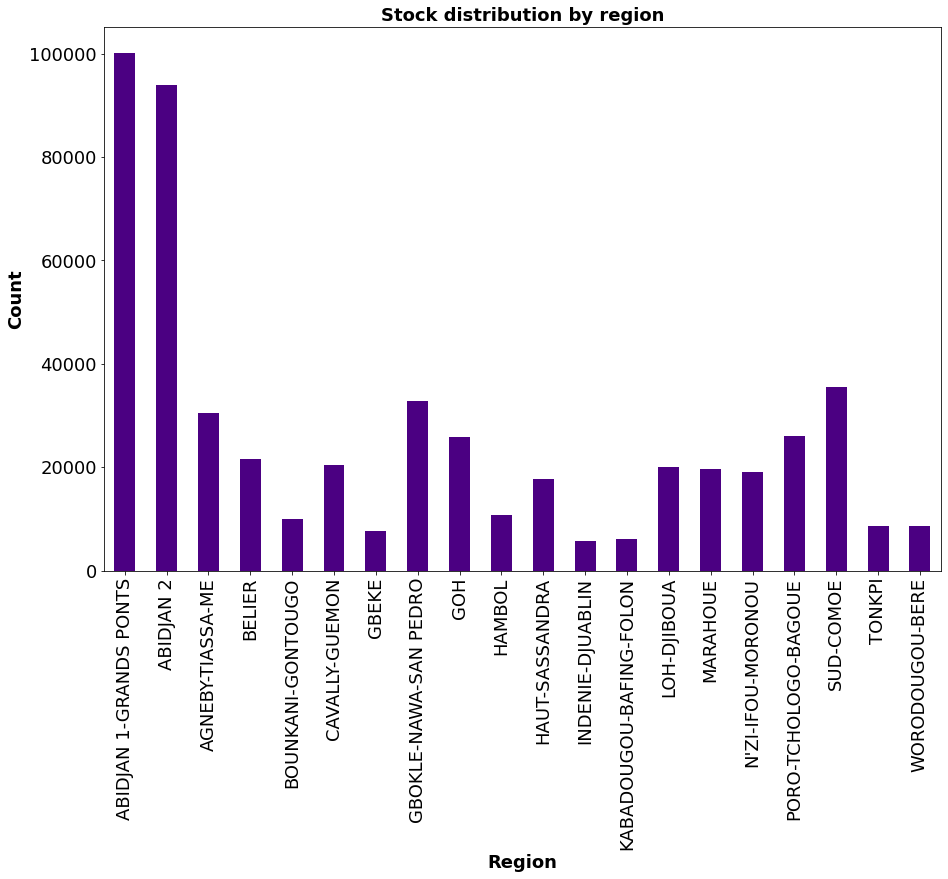
1. **District Distribution**



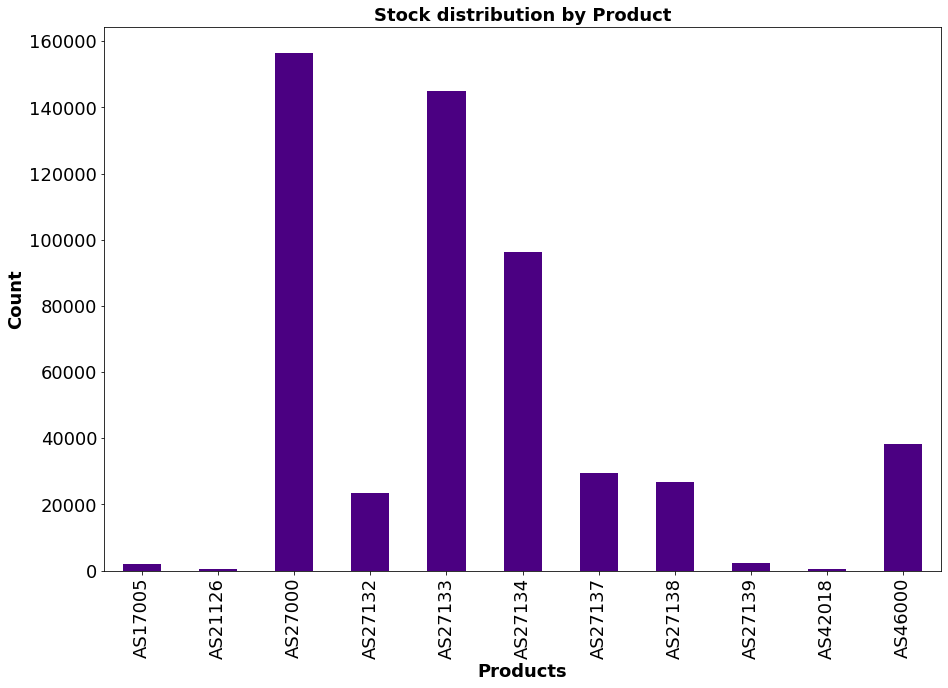
Koumassi-port bouet is the district that has the highest number of people using contraceptive.

Bangolo is the district that has the least number of people using contraceptive.

Stock Distribution by Region



Stock Distribution by Product

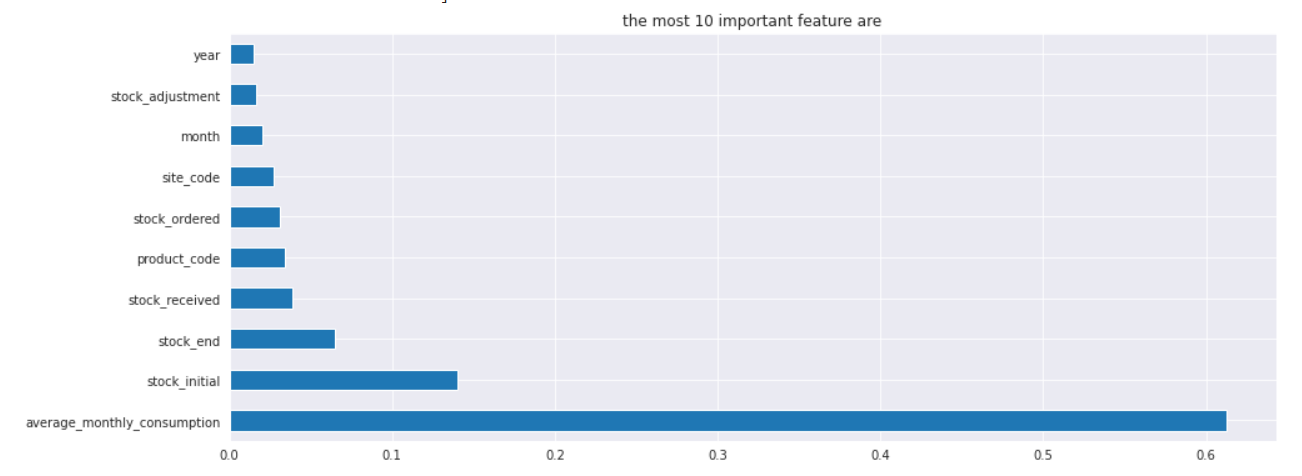


**5. MODELING AND EVALUATION**

Being a regression problem, the algorithms selected for building the model were Neural network, Catboost, XGboost and Random forest.

Before building the model, Feature selection was done using the Extratrees Regressor and Lasso Regression.

The 10 best features using **ExtraTrees** were as follows:



Using **Lasso Regression**, the important variables were as follows:

Stock Initial, Stock Received, Stock Adjustment, Stock End, Average Monthly Consumption and Stock ordered.

The ones that were discarded were: Year, Month, Site Code and Product Code. With this, the models were built using the features selected using Lasso. Performance of the models:

| **Model** | **RMSE** |
| --- | --- |
| Neural Network | 5.72 |
| CATBoost | 17.44 |
| XGBoost | 20.62 |
| RandomForest | 21.38 |

**6. CONCLUSIONS AND RECOMMENDATIONS**

**CONCLUSIONS**

* The most popular contraceptive was Injectable Contraceptive(DEPO PROVERA 150mg)
* The least popular contraceptive is the medroxtprogesterone
* The most visited site for contraceptives was the hospital
* The most prefered form of contraceptive was the injectable contraceptive and the least prefered female condom
* Koumassi-port bouet is the district that has the highest number of people using contraceptive while Bangolo is the district that has the least number of people using contraceptive.

**RECOMMENDATIONS**

* Ensure there is enough supply of contraceptives in regions with low supply like Bangolo and Agnibilekrou District
* Create awareness on the use of contraceptives in family planning methods,their use, benefits, and side effects
* They should implement Neural Networks for predicting the future use of contraceptives since it had the best RMSE compared to the other models