**Data Descriptions**

**Table 1: Products.txt**

* **Ingredient**  
  The active ingredient(s) for the product. Multiple ingredients are in alphabetical order, separated by a semicolon.

***My Notes:*** *maybe interesting to create a new column indicating the number of APIs that constitute each product. With this information we could know the* *mode/median/mean number of API in the composition of FDA-approved drugs. Is it different between innovators and generics? Is it different between Rx and OTC?*

*We could create a new Boolean column called Combination Drug to know whether the product is composed by one or more active ingredients. Then we could ask:* *What percentage represent combination drugs (multiple APIs in their composition) with respect to the total FDA-Approved product? Is it different between innovators and generics? Is it different between Rx and OTC?*

*Another question to answer would be…**Which are the most popular APIs in FDA-approved products’ composition? Is it different between innovators and generics? Is it different between Rx and OTC? For this aim we should somehow extract all the different APIs in a separate column and count them (value\_counts).*

* **Dosage form; Route of Administration**

The product dosage form and route separated by a semi-colon. The format is not all uppercase.

***My Notes:*** *maybe interesting to turn all values to uppercase and somehow separate dosage form and administration routes in two columns and then calculate* *which are the most common dosage form and administration routes in FDA-approved drugs. Is it different between innovators and generics?* *Is it different between Rx and OTC?*

* **Trade Name**

The trade name of the product as shown on the labelling.

* **Applicant**  
  The firm name holding legal responsibility for the new drug application. The firm name is condensed to a maximum twenty character unique string.

***My Notes:*** *which are the firms with the highest quantity of FDA-approved products (innovators/generics/Rx/OTC/)? And by Dosage Form? And by Route of administration?*

* **Strength**  
  The potency of the active ingredient. May repeat for multiple part products.
* **New Drug Application Type**

The type of new drug application approval. New Drug Applications (NDA or innovator) are ”N”. Abbreviated New Drug Applications (ANDA or generic) are “A”.

***My Notes:*** *interesting to substitute N by Innovator and A by Generic. This column is very useful for filtering.* *What percentage represent innovators and generics with respect to the total number of FDA-approved products? And within prescription drugs? Or within OTC? What percentage represents Rx and OTC within innovators? And within generics?*

* **New Drug Application (NDA) Number**

The FDA assigned number to the application. Format is nnnnnn

***My Notes:****How many different new drug applications (NDA or ANDAs) are registered at the FDA orange book at this moment*?

* **Product Number**

The FDA assigned number to identify the application products. Each strength is a separate product. May repeat for multiple part products. Format is nnn.

***My Notes:****For example, one product can have multiple strengths (Amoxicillin 100 mg or Amoxicillin 50 mg for the kids version). Different strengths have different product numbers, however, they constitute one single NDA or ANDA. One product may be constituted by different parts, for example, an inhaler (device + canister). Both parts share product number, they are considered just one product.*

* *How many FDA-approved products (drugs) are registered at the FDA Orange book at this moment? How many different products (product numbers) are normally included in a new drug application that has been approved by the FDA (mode, mean, median)? Is it different for innovators and generics? Is it different between Rx and OTC?*
* **Therapeutic Equivalence (TE) Code**

The TE Code indicates the therapeutic equivalence rating of generic to innovator Rx products.

**My Notes:** *Therapeutic equivalence codes are useful to indicate if the drug can be a substitute of the reference listed drug (RLD). Maybe we could answer…**What percentage of FDA-approved generics are considered therapeutically equivalent to their reference listed drug (RLD)? What percentage is not?*

* **Approval Date**

The date the product was approved as stated in the FDA approval letter to the applicant.  The format is Mmm dd, yyyy. Products approved prior to the January 1, 1982 contain the phrase: "Approved prior to Jan 1, 1982".

***My Notes:****Which has been the FDA approval tendency over the past years?*

* **Reference Listed Drug (RLD)**

The RLD is a drug product approved under section 505(c) of the FD&C Act for which FDA has made a finding of safety and effectiveness. In the electronic Orange Book, an RLD is identified by “RLD” in the RLD column.

* **Reference Standard (RS)**

A “reference standard” is the drug product selected by FDA that an applicant seeking approval of an ANDA must use in conducting an in vivo bioequivalence study required for approval of an ANDA. In the electronic Orange Book, a reference standard is identified by “RS” in the RS column.

**My Notes:** *What percentage of RLDs are considered RS by the FDA?*

* **Type**  
  The group or category of approved drugs. Format is RX, OTC, DISCN.

**My Notes:** *this field is a very good candidate for filtering, similar to Appl\_Type.* *Which firms have the highest numbers of FDA-approved prescription (Rx) drugs? What about OTC? And discontinued?*

* **Applicant Full Name**

The full name of the firm holding legal responsibility for the new drug application.

**Table 2: Patent.txt**

* **New Drug Application Type**

The type of new drug application approval.  New Drug Applications (NDA or innovator) are ”N”.  Abbreviated New Drug Applications (ANDA or generic) are “A”.

* **New Drug Application (NDA) Number**

The FDA assigned number to the application. Format is nnnnnn.

* **Product Number**

The FDA assigned number to identify the application products. Each strength is a separate product. May repeat for multiple part products. Format is nnn.

**My Notes:** *How many FDA-approved products (drugs) are registered at the FDA Orange book at this moment and have an associated patent? How many different products (product numbers) are normally included in a new drug application that has been approved by the FDA and have an associated patent (mode, mean, median)? Is it different for innovators and generics?*

* **Patent Number**

Patent numbers as submitted by the applicant holder for patents covered by the statutory provisions. May repeat for multiple applications and multiple products. Includes pediatric exclusivity granted by the agency. Format is nnnnnnnnnnn.

**My Notes:** *which is the total number of patents from FDA-approved drugs?*

* **Patent Expire Date**

The date the patent expires as submitted by the applicant holder including applicable extensions. The format is MMM DD, YYYY.

**My Notes:** *which has been the patent expiration tendency over the years?*

* **Drug Substance Flag**

Patents submitted on FDA Form 3542 and listed after August 18, 2003 may have a drug substance flag indicating the sponsor submitted the patent as claiming the drug substance.   Format is Y or null.

**My Notes:** *Not interesting for the analysis.*

* **Drug Product Flag**

Patents submitted on FDA Form 3542 and listed after August 18, 2003 may have a drug product flag indicating the sponsor submitted the patent as claiming the drug product.   Format is Y or null.

**My Notes:** *Not interesting for the analysis.*

* **Patent Use Code**

Code to designate a use patent that covers the approved indication or use of a drug product.  May repeat for multiple applications, multiple products and multiple patents. Format is nnnnnnnnnn.

**My Notes:** *we could extract information on the* *mean/median/mode number of different uses that are commonly claimed in a patent.*

* **Patent Delist Request Flag**

Sponsor has requested patent be delisted. This patent has remained listed because, under Section 505(j)(5)(D)(i) of the Act, a first applicant may retain eligibility for 180-day exclusivity based on a paragraph IV certification to this patent for a certain period. Applicants under Section 505(b)(2) are not required to certify to patents where this flag is set to Y. Format is Y or null.

**My Notes:** *What percentage of patents have been requested to be to delisted by the applicant? Which applicants have requested to delist patents the most (take into account the total number of patents per applicant)?*

* **Patent Submission Date**

The date on which the FDA receives patent information from the new drug application (NDA) holder. Format is Mmm d, yyyy

**My Notes:** *we could calculate the difference between patent expiration date and submission date in order to assess the* *mean/median/mode patent duration. Which has been the patent submission tendency throughout the years?*

**Table 3: Exclusivity.txt**

* **New Drug Application Type**

The type of new drug application approval. New Drug Applications (NDA or innovator) are ”N”. Abbreviated New Drug Applications (ANDA or generic) are “A”.

***My Notes:*** *which is the percentage of innovators and generics with respect to the total new drug applications with a granted exclusivity?*

* **New Drug Application (NDA) Number**

The FDA assigned number to the application. Format is nnnnnn.

***My Notes:****What’s the total number of new drug applications (NDA or ANDAs) with a granted FDA exclusivity?*

* **Product Number**

The FDA assigned number to identify the application products. Each strength is a separate product. May repeat for multiple part products. Format is nnn.

***My notes:*** *For example, one product can have multiple strengths (Amoxicillin 100 mg or Amoxicillin 50 mg for the kids version). Different strengths have different product numbers, however, they constitute one single NDA or ANDA. One product may be constituted by different parts, for example, an inhaler (device + canister). Both parts share product number, they are considered just one product.*

*How many different products (product numbers) are normally included in a new drug application that has been granted an FDA exclusivity (media/mode/median)? It is different for innovators or generics?*

* **Exclusivity Code**

Code to designate exclusivity granted by the FDA to a drug product. Format is nnnnnnnnnn.

***My Notes:*** *Each type of exclusivity has its own code (See Patents-and-Exclusivity-May-19--2015-Issue.pdf).*

*Which are the most popular post-approval exclusivities granted by the FDA? Do certain exclusivities tend to be granted together? (e.g., do some applications have multiple exclusivity types?)*

*Which exclusivity codes are found only in innovator (N) applications?*

* **Exclusivity Date**

The date the exclusivity expires. Format is MMM DD, YYYY.

***My Notes:****How many exclusivity periods expire each year in the dataset?*

*What percentage of new drug applications registered at the FDA Orange Book have been granted an FDA exclusivity? (for this I need those applications not appearing at the exclusivities table)*

**FDA Post-Approval Exclusivity Codes and other relevant information:**

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

El contenido generado por IA puede ser incorrecto.

When **exclusivity is granted** while the **patent is still active**, the NDA applicant benefits in several ways, as the exclusivity and patent provide **complementary protections**. Here's how:

**1️. Blocking Generic Competition More Effectively**

* **Exclusivity (FDA-based protection):**
  + Prevents the FDA from **approving or even reviewing** generic (ANDA) or 505(b)(2) applications for a set period.
  + Works **independently** of patents.
* **Patent (Legal protection):**
  + Prevents generics from **launching** the drug unless they successfully **challenge the patent** in court.
  + Allows the NDA holder to **sue** generic manufacturers for patent infringement.

✅ **Benefit:** If an NDA applicant holds both exclusivity and an active patent, **generics are blocked from both filling applications and launching the product, maximizing market control**.

**2. Delaying Generic Entry Even After Exclusivity Ends**

Even if **exclusivity expires**, a strong **patent portfolio** can still prevent generic entry.

🔹 **Example Scenario:**

* **Drug approval:** Jan 1, 2025
* **NCE Exclusivity (5 years) expires:** Jan 1, 2030
* **Patent expires:** Jan 1, 2040

**What happens?**

* Until **Jan 1, 2030** → **FDA cannot accept generic applications** due to **NCE exclusivity**.
* After **Jan 1, 2030** → Generics **can file** ANDAs, but if patents are still valid until **2040**, they must either:
  + Wait until **2040** to launch, OR
  + **Challenge the patents in court (Paragraph IV certification)**—which can take years.

✅ **Benefit:** Even after exclusivity ends, **patents extend market exclusivity for years, keeping generics out**.

**3. Extending Revenue Maximization**

* During the **exclusivity period**, the NDA holder enjoys **premium pricing** without generic competition.
* Even after exclusivity ends, **patents prevent generics from launching**, allowing continued revenue from **branded sales**.
* This is **especially useful if multiple exclusivities apply** (e.g., **NCE + Pediatric Exclusivity** adds 6 more months).

✅ **Benefit:** **More time to recoup R&D costs and maximize profits** before generic erosion starts.

**4. Patent Linkage in the Orange Book**

* NDA holders **list patents in the FDA's Orange Book**.
* If a generic company **files a Paragraph IV certification** challenging a listed patent, the NDA holder gets an **automatic 30-month stay** of FDA approval while the patent lawsuit is resolved.

✅ **Benefit:** **Further delays generic competition** even after exclusivity ends

**🔹 Conclusion:**

When **both exclusivity and patents overlap**, the NDA holder benefits from **multiple layers of protection**, keeping generics out for as long as possible and **maximizing market exclusivity & profits**. 🚀