

## 零 项目流程

### 0.1 数据流程

### 0.2 代码流程

## 一 业务理解

### 1.1关于NFL

## 二 数据理解

### 2.1 字段和键定义

#### 2.1.1 Injury Data File

#### 2.1.2 Play List

#### 2.1.3 Player Track Data

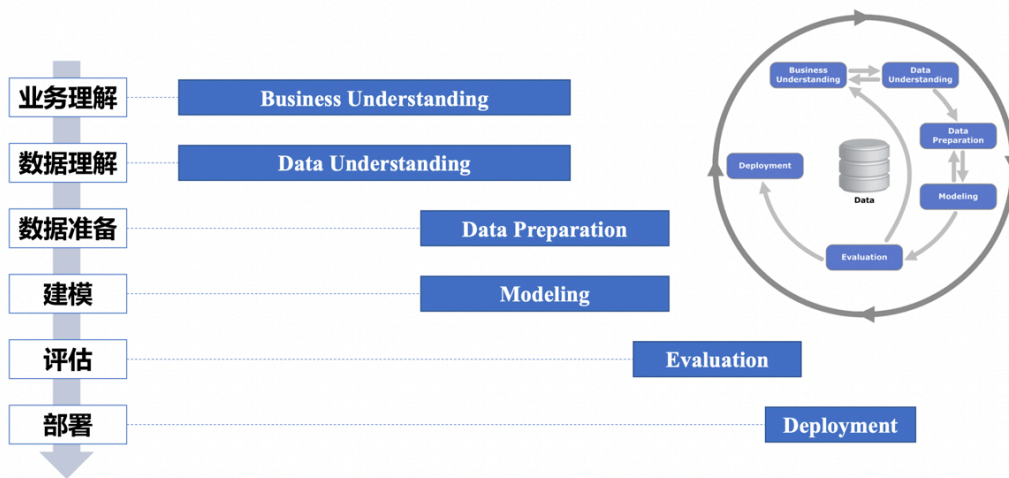
## 三 数据准备

## 四 项目进度自查

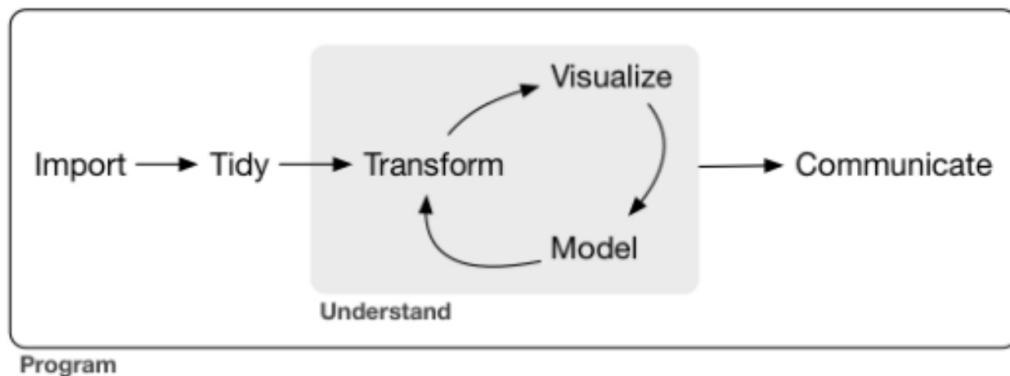
# 零 项目流程

## 0.1 数据流程

## 数据挖掘的一般流程



### 0.2 代码流程



## 一 业务理解

### 1.1 关于NFL

美国国家橄榄球联盟（National Football League）是美国最受欢迎的体育联盟，由32个专营权组成的联盟每年竞争赢得超级碗，这是世界上最大的年度体育赛事。NFL成立于1920年，为成功的现代体育联赛开发了该模式，包括国内和国际发行，广泛的收益共享，卓越的竞争以及遍布全国的强大特许经营权。

NFL致力于在运动相关伤害的诊断，预防和治疗方面取得进步。NFL正在进行的健康与安全工作包括对独立医学研究和工程进展的支持，以及致力于更好地保护玩家并使游戏更加安全的承诺，包括增强医疗协议以及改进我们的游戏教学和玩法。

从中学到的更多信息，联盟会评估和更改规则以发展比赛并尝试改善对球员的保护。仅从2002

年以来，美国国家橄榄球联盟就进行了50项规则更改，旨在消除潜在的危险战术并减少受伤风险。

## 二 数据理解

数据集中提供了三个文件，如下所述：

- Injury Record: .csv格式的伤害记录文件包含有关在两个赛季的常规赛期间发生的105下肢伤害的信息。可以使用PlayerKey，GameID和PlayKey字段将伤害链接到玩家历史记录中的特定记录。
- Play List: 有关比赛和比赛的详细信息包括球员分配的名册位置，体育场类型，场地类型，天气，比赛类型，比赛位置和位置组。
- Player Track Data:用于描述以10 Hz记录的播放过程中每个玩家的位置，方向，速度和方向（即，每秒记录10个观测值）。

### 2.1 字段和键定义

数据库	描述	球员键	比赛ID	发挥键
受伤记录	关于受伤的信息	x	x	x
比赛文件	比赛、场地、球员-比赛水平的发挥信息	x	x	x
球员追踪	球员位置、速度、在一场比赛中球员个人的方向信息			x

下面提供了数据集中包含的每个字段的描述及其相应的格式和描述。主键 外键 变量以粗体指定。

#### 2.1.1 Injury Data File

字段	格式	描述
球员关键 (PlayerKey)	XXXX	用五位数的数字键标识一位球员
比赛ID (Game ID)	PlayerKey-X	只识别球员的游戏 (不严格按照时间顺序)
发挥关键 (PlayKey)	PlayerKey-GameID-X	只识别比赛中球员的发挥 (按顺序)
身体部位 (BodyPart)	字符串	识别受伤的身体部位 (膝盖、脚踝、脚等……)
场地 (Surface)	字符串	识别受伤时的比赛场地 (天然的还是人造的)
DM_M1	1或0	One-hot编码, 指示由于受伤错过了1天或更多天
DM_M7	1或0	One-hot编码, 指示由于受伤错过了7天或更多天
DM_M28	1或0	One-hot编码, 指示由于受伤错过了28天或更多天
DM_M42	1或0	One-hot编码, 指示由于受伤错过了42天或更多天

### 2.1.2 Play List

字段	格式	描述
球员关键 (PlayerKey)	XXXX	用五位数的数字键标识一位球员
比赛ID (Game ID)	PlayerKey-X	只识别球员-比赛 (该索引不严格按照时间顺序)-请参阅下面的注释
发挥关键 (PlayKey)	PlayerKey-GameID-X	只识别比赛中球员的发挥 (按比赛顺序)
名册位置(RosterPosition)	字符串	提供球员的名册位置 (即返回)
*比赛时间(PlayerDay)	整型	一个整型序列, 它反映了球员在比赛中的参与时间线; 使用这个来按球员参与的顺序。
*球员比赛 (PlayerGame)	整型	只识别一个球员的比赛; 匹配比赛ID的最后一个整型 (并不严格按照比赛的时间顺序)
体育场类型(StadiumType)	字符串	体育馆类型的自由文字说明 (闭式、开放式圆顶等等)
场地类型(FiledType)	字符串	场地种类的分类描述 (自然的还是人造的)
温度(Temperature)	浮点数	比赛开始时的现场温度 (并不总是可用的-对于闭式圆顶/户内体育场于受温度和天气的影响, 此项可能是不相关的)
天气 (Weather)	字符串	描述体育场天气的自由文字说明 (由于受温度和天气的影响, 此项可能是不相关的)
玩法类型 (PlayType)	字符串	玩法的分类描述 (传球、跑、踢等等)
PlayerGamePlay		指示球员在比赛过程中参与次数的顺序索引
位置 (Position)	字符串	指示球员在比赛中的位置的分类变量 (RB,QB,DE等) -也许和名册位置不同
队伍位置 (Position Group)	字符串	表示球员在比赛中所处位置的位置组的分类变量

### 2.1.3 Player Track Data

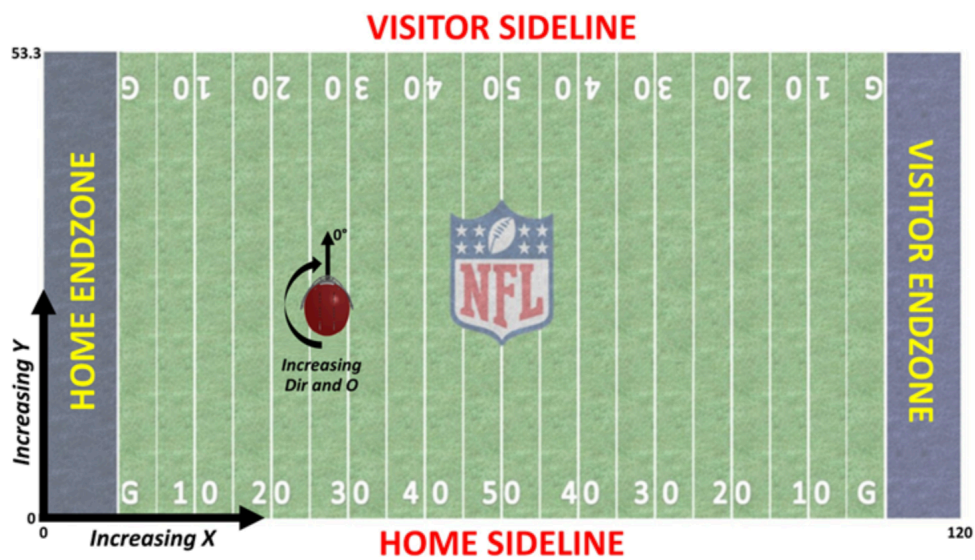
字段	格式	描述	补充细节
发挥关键(PlayKey)	字符串		
时间 (time)	浮点型	从该比赛的NGS轨道开始的秒数（浮点数）	这是运动员轨迹的时间
事件 (event)	字符串	随时间变化的比赛细节（字符串）（huddle break,snap等等）	
x	数字列表	沿着场地的长轴随时间变化的球员位置	0~120码（yards）
y	数字列表	沿着场地的短轴随时间变化的球员位置	0~53.3码（yards）
dis	数字列表	从先前时间点到时间索引的行进距离	距离（码（yards））
s	数字列表	时间索引上该特定时间点的估计速度	码每秒
o	数字	方向-球员面朝的角度（deg）	0~360度
dir	数字	方向-球员运动的角度（deg）	0~360度

处理玩家轨迹数据时，建议使用x，y位置数据计算速度，并将这些计算出的速度用于任何分析（尽管我们提供了NGS系统报告的速度变量）。x和y坐标的原点定义为起始端区和起始边线的角（见图1）。由方向和方向定义的角度是从坐标系的y轴参考的。

注意，方向变量不应被视为玩家面对的实际方向的可靠指标。这项研究的记录来自NFL的多个赛季，在此期间，使用了不同的系统来计算和记录球员的方位。在一个游戏内以及游戏中的多个游戏中，单个玩家的方向将被一致地记录，但是报告的方向的“地理位置”（即玩家所面对的实际方向）可能与方向变量的“地理位置”不匹配在同一戏剧中。对于参与多个季节的那些球员，用于记录方向变量的地理位置在各个季节之间都不相同。方向变量可用于表征玩家在比赛过程中转动或旋转多少。

### 三 数据准备

以提供



## 四 项目进度自查

- ☒ Import 包环境依赖,导入数据
- ☒ Tidy 清洗数据(错别字/空字段)
- ☐ Transform数据转换/翻译工作
- ☐ vis 数据可视化
- ☐ model 数据建模
- ☐ com 评估工作